

**This Page Is Inserted by IFW Operations  
and is not a part of the Official Record**

## **BEST AVAILABLE IMAGES**

**Defective images within this document are accurate representations of the original documents submitted by the applicant.**

**Defects in the images may include (but are not limited to):**

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:30:32 ; Search time 8.35496 Seconds  
(without alignments)  
1097.829 Million cell updates/sec

Title: US-09-816-653A-2  
Perfect score: 1061  
Sequence: 1 MSQAPNGTSPGATEDYSYG.....MAHLGVQVWRAECPPVKI 199

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEM\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC7\_NEM\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEM\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEM\_PUB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09\_NEM\_PUB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US10\_NEM\_PUB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US60\_NEM\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1059	99.8	199	US-09-816-653A-2	Sequence 2, Appl1
2	1048	98.8	667	US-09-759-056-2	Sequence 2, Appl1
3	1048	98.8	667	US-09-901-812-2	Sequence 80, Appl1
4	1048	98.8	667	US-10-227-884-80	Sequence 80, Appl1
5	1048	98.8	667	US-10-230-163-80	Sequence 80, Appl1
6	1048	98.8	667	US-10-218-631-80	Sequence 80, Appl1
7	1048	98.8	667	US-10-230-338-80	Sequence 80, Appl1
8	986.5	93.0	658	US-09-759-056-5	Sequence 5, Appl1
9	986.5	93.0	658	US-09-901-812-5	Sequence 5, Appl1
10	697	65.7	670	US-09-816-653A-7	Sequence 7, Appl1
11	84	7.9	624	US-09-925-297-738	Sequence 738, App
12	83	7.8	913	US-09-223-490-4	Sequence 4, Appl1
13	78.5	7.4	3519	US-09-868-880-4	Sequence 4, Appl1
14	74.5	7.0	208	US-09-738-626-6358	Sequence 6358, Ap
15	74	7.0	328	US-09-925-301-1349	Sequence 1349, Ap
16	73.5	6.9	356	US-09-738-626-4761	Sequence 4761, Ap
17	73.5	6.9	378	US-10-021-756-2	Sequence 2, Appl1
18	73.5	6.9	636	US-10-000-776-12	Sequence 12, Appl1
19	73.5	6.9	636	US-09-791-497-14	Sequence 14, Appl1

20	73	6.9	633	US-10-086-464-11	Sequence 11, Appl1
21	72.5	6.8	504	US-09-758-386-2	Sequence 2, Appl1
22	72.5	6.8	504	US-09-815-108-17	Sequence 17, Appl1
23	72.5	6.8	504	US-09-815-108-17	Sequence 17, Appl1
24	72.5	6.8	529	US-09-712-363-175	Sequence 175, App
25	72	6.8	183	US-09-764-853-822	Sequence 822, App
26	72	6.8	652	US-10-139-662-4	Sequence 4, Appl1
27	72	6.8	694	US-09-846-996A-1	Sequence 1, Appl1
28	70.5	6.6	627	US-09-853-386-67	Sequence 67, Appl1
29	70.5	6.6	627	US-10-005-169-2	Sequence 2, Appl1
30	70.5	6.6	628	US-09-853-386-66	Sequence 66, Appl1
31	70.5	6.6	628	US-09-853-386-63	Sequence 63, Appl1
32	70.5	6.6	628	US-09-853-386-73	Sequence 73, Appl1
33	70.5	6.6	628	US-10-005-169-4	Sequence 4, Appl1
34	70.5	6.6	730	US-09-738-626-4196	Sequence 4196, Ap
35	69.5	6.6	206	US-09-800-729-149	Sequence 149, App
36	69.5	6.6	369	US-10-077-874-4	Sequence 4, Appl1
37	69.5	6.6	372	US-09-393-696-6	Sequence 6, Appl1
38	69.5	6.6	372	US-10-077-874-6	Sequence 6, Appl1
39	69.5	6.6	389	US-09-828-538-20	Sequence 20, Appl1
40	69.5	6.6	425	US-09-828-538-19	Sequence 19, Appl1
41	69.5	6.6	425	US-09-828-538-24	Sequence 24, Appl1
42	69.5	6.6	425	US-10-139-662-2	Sequence 2, Appl1
43	69	6.5	122	US-09-796-858-14	Sequence 14, Appl1
44	69	6.5	238	US-10-001-054-34	Sequence 34, Appl1
45	69	6.5	238	US-10-006-856A-271	Sequence 271, App

#### ALIGNMENTS

RESULT 1  
US-09-816-653A-2  
; Sequence 2, Application US/0981653A  
; Publication No. US20030021788A1  
; GENERAL INFORMATION:  
; APPLICANT: Rastelli, Diane  
; TITLE OF INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME  
; FILE REFERENCE: 10716/57  
; CURRENT APPLICATION NUMBER: US/09/816,653A  
; CURRENT FILING DATE: 2001-03-23  
; PRIOR APPLICATION NUMBER: 60/191,532  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 2  
; LENGTH: 199  
; TYPE: PRT  
; ORGANISM: Homo sapiens hSTRa6 amino terminal polypeptide fragment  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (31)-(31)  
; OTHER INFORMATION: Xaa is any amino acid  
US-09-816-653A-2

Query Match 99.8%; Score 1059; DB 9; Length 199;  
Best Local Similarity 100.0%; Pred. No. 1.3e-95;  
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSQAPNGTSPGATEDYSYGWYIDEPQGXELQPEGEVPSCHTSIPPGIYHACLASLSI 60  
DB 1 MSQAPNGTSPGATEDYSYGWYIDEPQGXELQPEGEVPSCHTSIPPGIYHACLASLSI 60  
QY 61 LVLLLLMLVRRRLQIPDCVRRRLPSVPVPLAGDRRAVPAVFWVLSSICLLLPDE 120  
DB 61 LVLLLLMLVRRRLQIPDCVRRRLPSVPVPLAGDRRAVPAVFWVLSSICLLLPDE 120  
QY 121 DALPFLTLAASPQDQKTEAPRGAMWIKLGFHYALYYPLAACATGATTAHLLSGTSLW 180  
DB 121 DALPFLTLAASPQDQKTEAPRGAMWIKLGFHYALYYPLAACATGATTAHLLSGTSLW 180  
QY 181 AHLGVQVWRAECPPVKI 199

Db 181 HLGVQVWQRAECPQVPK1 199

## RESULT 2

US-09-759-056-2

Sequence 2, Application US/09759056

Patent No. US20020156252A1

GENERAL INFORMATION:

APPLICANT: Pennica, Diane

APPLICANT: Smith, Victoria

APPLICANT: Wood, William I.

TITLE OF INVENTION: NO. US20020156252A1 STRA6 Polypeptides

FILE REFERENCE: GENENT.2827A2

CURRENT APPLICATION NUMBER: US/09/759,056

CURRENT FILING DATE: 2001-01-11

PRIOR APPLICATION NUMBER: 60/197089

PRIOR FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/175849

PRIOR FILING DATE: 2000-01-13

PRIOR APPLICATION NUMBER: 60/228914

PRIOR FILING DATE: 2000-08-29

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 667

TYPE: PRT

ORGANISM: Homo Sapiens

US-09-759-056-2

Query Match 98.8%; Score 1048; DB 9; Length 667;

Best Local Similarity 99.0%; Pred. No. 6,5e-94;

Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLSTL 61  
DB 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLSTL 62  
QY 62 VLLILAMLVRRRLQMPDCCVGRGRLPSVDFLAGDPRAPVPAVFWLLSSLCILLPDED 121  
DB 63 VLLILAMLVRRRLQMPDCCVGRGRLPSVDFLAGDPRAPVPAVFWLLSSLCILLPDED 122  
QY 122 ALPFLTLASAPSDGKTEAPRGAMKILGFHYAALYPLAACATAGHTAAHLGSLTSLWA 181  
DB 123 ALPFLTLASAPSDGKTEAPRGAMKILGFHYAALYPLAACATAGHTAAHLGSLTSLWA 182  
QY 182 HLGVQVWQRAECPQVPK1 199  
DB 183 HLGVQVWQRAECPQVPK1 200

## RESULT 3

US-09-901-812-2

Sequence 2, Application US/09901812

Patent No. US20020173461A1

GENERAL INFORMATION:

APPLICANT: Pennica, Diane

APPLICANT: Polakis, Paul

APPLICANT: Szeto, Wayne

TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO

FILE REFERENCE: GENENT.083A

CURRENT APPLICATION NUMBER: US/09/901,812

CURRENT FILING DATE: 2001-07-10

PRIOR APPLICATION NUMBER: 60/228,914

PRIOR FILING DATE: 2000-08-29

PRIOR APPLICATION NUMBER: 09/759,056

PRIOR FILING DATE: 2001-01-11

PRIOR APPLICATION NUMBER: 60/175,849

PRIOR FILING DATE: 2000-01-13

PRIOR APPLICATION NUMBER: 60/197,089

PRIOR FILING DATE: 2000-04-14

NUMBER OF SEQ ID NOS: 28

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 667  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-901-812-2

Query Match 98.8%; Score 1048; DB 9; Length 667;

Best Local Similarity 99.0%; Pred. No. 6,5e-94;

Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLSTL 61  
DB 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLSTL 62  
QY 62 VLLILAMLVRRRLQMPDCCVGRGRLPSVDFLAGDPRAPVPAVFWLLSSLCILLPDED 121  
DB 63 VLLILAMLVRRRLQMPDCCVGRGRLPSVDFLAGDPRAPVPAVFWLLSSLCILLPDED 122  
QY 122 ALPFLTLASAPSDGKTEAPRGAMKILGFHYAALYPLAACATAGHTAAHLGSLTSLWA 181  
DB 123 ALPFLTLASAPSDGKTEAPRGAMKILGFHYAALYPLAACATAGHTAAHLGSLTSLWA 182  
QY 182 HLGVQVWQRAECPQVPK1 199  
DB 183 HLGVQVWQRAECPQVPK1 200

## RESULT 4

US-10-227-884-80

Sequence 80, Application US/10227884

Publication No. US20030027988A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Desnoyers, Luc

APPLICANT: Gerritsen, Mary

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, J. Paul J.

APPLICANT: Grimaldi, J. Christopher

APPLICANT: Gurney, Austin L.

APPLICANT: Smith, Victoria

APPLICANT: Stephan, Jean-Philippe F.

APPLICANT: Watanabe, Colin L.

APPLICANT: Wood, William I.

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3530P1C79

CURRENT APPLICATION NUMBER: US/10/227,884

CURRENT FILING DATE: 2002-08-26

PRIOR APPLICATION NUMBER: 10/119,480

PRIOR FILING DATE: 2002-04-09

PRIOR APPLICATION NUMBER: 60/059113

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/062287

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/063549

PRIOR FILING DATE: 1997-10-28

PRIOR APPLICATION NUMBER: 60/064103

PRIOR FILING DATE: 1997-10-31

PRIOR APPLICATION NUMBER: 60/069873

PRIOR FILING DATE: 1997-12-17

PRIOR APPLICATION NUMBER: 60/078910

PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/079294

PRIOR FILING DATE: 1998-03-25

PRIOR APPLICATION NUMBER: 60/079656

PRIOR FILING DATE: 1998-03-26

PRIOR APPLICATION NUMBER: 60/079728

PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/081819

PRIOR FILING DATE: 1998-04-15

PRIOR APPLICATION NUMBER: 60/081955

PRIOR FILING DATE: 1998-04-15



PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089905  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090691  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/095302  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/095318  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/095916  
PRIOR FILING DATE: 1998-08-10  
PRIOR APPLICATION NUMBER: 60/096146  
PRIOR FILING DATE: 1998-08-11  
PRIOR APPLICATION NUMBER: 60/096791  
PRIOR FILING DATE: 1998-08-17  
PRIOR APPLICATION NUMBER: 60/097986  
PRIOR FILING DATE: 1998-08-26  
PRIOR APPLICATION NUMBER: 60/098544  
PRIOR FILING DATE: 1998-08-31  
PRIOR APPLICATION NUMBER: 60/099596  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099803  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099811  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099812  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099816  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100038  
PRIOR FILING DATE: 1998-09-11  
PRIOR APPLICATION NUMBER: 60/100385  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100627  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100848  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/100919  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/101477  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101738  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101741  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101786  
PRIOR FILING DATE: 1998-09-25  
PRIOR APPLICATION NUMBER: 60/101916  
  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101922  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/106178  
PRIOR FILING DATE: 1998-10-28  
PRIOR APPLICATION NUMBER: 60/106248  
PRIOR FILING DATE: 1998-10-29  
PRIOR APPLICATION NUMBER: 60/106464  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: 60/106905  
PRIOR FILING DATE: 1998-11-03  
PRIOR APPLICATION NUMBER: 60/108787  
PRIOR FILING DATE: 1998-11-17  
PRIOR APPLICATION NUMBER: 60/108801  
PRIOR FILING DATE: 1998-11-17  
PRIOR APPLICATION NUMBER: 60/108849  
PRIOR FILING DATE: 1998-11-18  
PRIOR APPLICATION NUMBER: 60/112422  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113296  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/113605  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113621  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/115558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119549  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/123618  
PRIOR FILING DATE: 1999-03-10  
PRIOR APPLICATION NUMBER: 60/125259  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 60/125775  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/126773  
PRIOR FILING DATE: 1999-03-29  
PRIOR APPLICATION NUMBER: 60/127887  
PRIOR FILING DATE: 1999-04-05  
PRIOR APPLICATION NUMBER: 60/130232  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/131022  
PRIOR FILING DATE: 1999-04-26  
PRIOR APPLICATION NUMBER: 60/131270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131445  
PRIOR FILING DATE: 1999-04-28  
PRIOR APPLICATION NUMBER: 60/134287  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/140650  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/140723  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/141037  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/144758  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/145698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 60/146222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: 60/146963  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/149320  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/149638  
PRIOR FILING DATE: 1999-08-17

```

PRIOR APPLICATION NUMBER: 60/151733
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/164418
PRIOR FILING DATE: 1999-11-09
PRIOR APPLICATION NUMBER: 60/166361
PRIOR FILING DATE: 1999-11-16
PRIOR APPLICATION NUMBER: 60/169445
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/169835

Query Match 98.8%; Score 1048; DB %; Length 667;
Best Local Similarity 99.0%; Pred. No. 6,5e-94;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPANQTSFGATEDYSGSMYIDEPQGGXELQPEGEVSCSTSPRGVYACLASLSTL 61
DB 3 SOPANQTSFGATEDYSGSMYIDEPQGGXELQPEGEVSCSTSPRGVYACLASLSTL 62
QY 62 VLLIAMLVRRRQIMPCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSICLLPDED 121
DB 63 VLLIAMLVRRRQIMPCVGRPGLPSPVDFLAGDRPRAVPAVFMVLLSSICLLPDED 122
QY 122 ALPFLTLASPSQDKTEAPRGAMKTLGLPHYALYPPAACATGHTAAHLGTSMA 161
DB 123 ALPFLTLASPSQDKTEAPRGAMKTLGLFYALYPPAACATGHTAAHLGTSMA 162
QY 162 HLGVQWQORACCPQPKI 199
DB 163 HLGVQWQORACCPQPKI 200

RESULT 5
US-10-230-163-80
Sequence 80. Application US/10230163
Publication No US2003003653A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Desnoyers, Luc
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Auerin L.
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Metanabe, Colin L.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3530P1C96
CURRENT APPLICATION NUMBER: US/10/230.163
PRIOR FILING DATE: 2002-08-28
PRIOR APPLICATION NUMBER: 10/119,480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/068873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079728
```

```

PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089905
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090691
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/095302
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095318
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095916
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/096146
PRIOR FILING DATE: 1998-08-11
PRIOR APPLICATION NUMBER: 60/096791
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/097986
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098544
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100038
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
```

PRIOR APPLICATION NUMBER: 60/101741  
 PRIOR FILING DATE: 1998-09-24  
 PRIOR APPLICATION NUMBER: 60/101786  
 PRIOR FILING DATE: 1998-09-25  
 PRIOR APPLICATION NUMBER: 60/101916  
 PRIOR FILING DATE: 1998-09-24  
 PRIOR APPLICATION NUMBER: 60/101922  
 PRIOR FILING DATE: 1998-09-24  
 PRIOR APPLICATION NUMBER: 60/106178  
 PRIOR FILING DATE: 1998-10-28  
 PRIOR APPLICATION NUMBER: 60/106248  
 PRIOR FILING DATE: 1998-10-29  
 PRIOR APPLICATION NUMBER: 60/106464  
 PRIOR FILING DATE: 1998-10-30  
 PRIOR APPLICATION NUMBER: 60/106905  
 PRIOR FILING DATE: 1998-11-03  
 PRIOR APPLICATION NUMBER: 60/108787  
 PRIOR FILING DATE: 1998-11-17  
 PRIOR APPLICATION NUMBER: 60/108801  
 PRIOR FILING DATE: 1998-11-17  
 PRIOR APPLICATION NUMBER: 60/108849  
 PRIOR FILING DATE: 1998-11-18  
 PRIOR APPLICATION NUMBER: 60/112422  
 PRIOR FILING DATE: 1998-12-15  
 PRIOR APPLICATION NUMBER: 60/113296  
 PRIOR FILING DATE: 1998-12-22  
 PRIOR APPLICATION NUMBER: 60/113605  
 PRIOR FILING DATE: 1998-12-23  
 PRIOR APPLICATION NUMBER: 60/113621  
 PRIOR FILING DATE: 1998-12-23  
 PRIOR APPLICATION NUMBER: 60/115558  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115565  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/115733  
 PRIOR FILING DATE: 1999-01-12  
 PRIOR APPLICATION NUMBER: 60/119549  
 PRIOR FILING DATE: 1999-02-10  
 PRIOR APPLICATION NUMBER: 60/123618  
 PRIOR FILING DATE: 1999-03-10  
 PRIOR APPLICATION NUMBER: 60/125259  
 PRIOR FILING DATE: 1999-03-19  
 PRIOR APPLICATION NUMBER: 60/125775  
 PRIOR FILING DATE: 1999-03-23  
 PRIOR APPLICATION NUMBER: 60/126773  
 PRIOR FILING DATE: 1999-03-29  
 PRIOR APPLICATION NUMBER: 60/127887  
 PRIOR FILING DATE: 1999-04-05  
 PRIOR APPLICATION NUMBER: 60/130232  
 PRIOR FILING DATE: 1999-04-21  
 PRIOR APPLICATION NUMBER: 60/131022  
 PRIOR FILING DATE: 1999-04-26  
 PRIOR APPLICATION NUMBER: 60/131270  
 PRIOR FILING DATE: 1999-04-27  
 PRIOR APPLICATION NUMBER: 60/131291  
 PRIOR FILING DATE: 1999-04-27  
 PRIOR APPLICATION NUMBER: 60/131445  
 PRIOR FILING DATE: 1999-04-28  
 PRIOR APPLICATION NUMBER: 60/134287  
 PRIOR FILING DATE: 1999-05-14  
 PRIOR APPLICATION NUMBER: 60/140650  
 PRIOR FILING DATE: 1999-06-22  
 PRIOR APPLICATION NUMBER: 60/140723  
 PRIOR FILING DATE: 1999-06-22  
 PRIOR APPLICATION NUMBER: 60/141037  
 PRIOR FILING DATE: 1999-06-23  
 PRIOR APPLICATION NUMBER: 60/144758  
 PRIOR FILING DATE: 1999-07-20  
 PRIOR APPLICATION NUMBER: 60/145698  
 PRIOR FILING DATE: 1999-07-26  
 PRIOR APPLICATION NUMBER: 60/146222  
 PRIOR FILING DATE: 1999-07-28  
 PRIOR APPLICATION NUMBER: 60/146963

PRIOR FILING DATE: 1999-08-03  
 PRIOR APPLICATION NUMBER: 60/149320  
 PRIOR FILING DATE: 1999-08-17  
 PRIOR APPLICATION NUMBER: 60/149638  
 PRIOR FILING DATE: 1999-08-17  
 PRIOR APPLICATION NUMBER: 60/151733  
 PRIOR FILING DATE: 1999-08-31  
 PRIOR APPLICATION NUMBER: 60/164418  
 PRIOR FILING DATE: 1999-11-09  
 PRIOR APPLICATION NUMBER: 60/166361  
 PRIOR FILING DATE: 1999-11-16  
 PRIOR APPLICATION NUMBER: 60/169445  
 PRIOR FILING DATE: 1999-12-07  
 PRIOR APPLICATION NUMBER: 60/169495  
 PRIOR FILING DATE: 1999-12-07  
 PRIOR APPLICATION NUMBER: 60/169835

Query Match 98.8%; Score 1048; DB 9; Length 667;  
 Best Local Similarity 99.0%; Pred. No. 6-5e-94;  
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSPGATEDYISGMYIDEPQGXELQPBGEVPSCHTSPGLYHACIASISIL 61  
 DB 3 SOPAGNOTSPGATEDYISGMYIDEPQGXELQPBGEVPSCHTSPGLYHACIASISIL 62  
 QY 62 VLLIAMLYRRRLMPDVCVGRPCLSPVDFLAGDRPRAVPAVEMVLLSSICLLPDED 121  
 DB 63 VLLIAMLYRRRLMPDVCVGRPCLSPVDFLAGDRPRAVPAVEMVLLSSICLLPDED 122  
 QY 122 ALPFTLASPSOGKTBPFGANKILGLFHYALYPLAACTGCHRAHLSTLSMA 181  
 DB 123 ALPFTLASPSOGKTBPFGANKILGLFHYALYPLAACTGCHRAHLSTLSMA 182  
 QY 182 HLGQVWQRAECQVPK1 199  
 DB 183 HLGQVWQRAECQVPK1 200

RESULT 6  
 US-10-218-631-80  
 Sequence 80, Application US/10218631  
 Publication No. US20030045687A1  
 GENERAL INFORMATION:  
 APPLICANT: Baker, Kevin P.  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Gerritsen, Mary  
 APPLICANT: Goddard, Audrey J.  
 APPLICANT: Godowski, Paul J.  
 APPLICANT: Grimaldi, J. Christopher  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Smith, Victoria  
 APPLICANT: Stephan, Jean-Philippe F.  
 APPLICANT: Watanabe, Colin L.  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 FILE REFERENCE: P3530PIC14  
 CURRENT APPLICATION NUMBER: US/10/218,631  
 PRIOR FILING DATE: 2002-08-12  
 PRIOR APPLICATION NUMBER: 10/119,480  
 PRIOR FILING DATE: 2002-04-09  
 PRIOR APPLICATION NUMBER: 60/059113  
 PRIOR FILING DATE: 1997-09-17  
 PRIOR APPLICATION NUMBER: 60/062287  
 PRIOR FILING DATE: 1997-10-17  
 PRIOR APPLICATION NUMBER: 60/063549  
 PRIOR FILING DATE: 1997-10-28  
 PRIOR APPLICATION NUMBER: 60/064103  
 PRIOR FILING DATE: 1997-10-31  
 PRIOR APPLICATION NUMBER: 60/069873  
 PRIOR FILING DATE: 1997-12-17  
 PRIOR APPLICATION NUMBER: 60/078910  
 PRIOR FILING DATE: 1998-03-20

```
/ PRIOR APPLICATION NUMBER: 60/079294
/ PRIOR FILING DATE: 1998-03-26
/ PRIOR APPLICATION NUMBER: 60/079656
/ PRIOR FILING DATE: 1998-03-26
/ PRIOR APPLICATION NUMBER: 60/079728
/ PRIOR FILING DATE: 1998-03-27
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 246
/ SEQ ID NO 80
/ LENGTH: 667
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-218-631-80

Query Match          98.8%; Score 1048; DB 9; Length 667;
Best Local Similarity 99.0%; Pred. No. 6.5e-94;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLIL 61
Db 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLIL 62
Qy 62 VLLILAMLVRRRLQMPDCVGRGRLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 121
Db 63 VLLILAMLVRRRLQMPDCVGRGRLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 122
Qy 122 ALPFTLLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSLTSLWA 181
Db 123 ALPFTLLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSLTSLWA 182
Qy 182 HLGVOVWQRAECPOVPKI 199
Db 183 HLGVOVWQRAECPOVPKI 200

RESULT 7
US-10-230-338-80
/ Sequence 80. Application US/10230338
/ Patent No. US20030044934A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Deenoyere, Luc
/ APPLICANT: Geriltsen, Mary
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Goddard, Paul J.
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe F.
/ APPLICANT: Macanabe, Colin L.
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3530PIC92
/ CURRENT APPLICATION NUMBER: US/10/230,338
/ PRIOR APPLICATION NUMBER: 2002-08-28
/ PRIOR FILING DATE: 2002-04-09
/ PRIOR APPLICATION NUMBER: 60/059113
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/062287
/ PRIOR FILING DATE: 1997-10-17
/ PRIOR APPLICATION NUMBER: 60/063549
/ PRIOR FILING DATE: 1997-10-28
/ PRIOR APPLICATION NUMBER: 60/064103
/ PRIOR FILING DATE: 1997-10-31
/ PRIOR APPLICATION NUMBER: 60/069873
/ PRIOR FILING DATE: 1997-12-17
/ PRIOR APPLICATION NUMBER: 60/078910
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/079294
/ PRIOR FILING DATE: 1998-03-25
/ PRIOR APPLICATION NUMBER: 60/079656
```

```
/ PRIOR FILING DATE: 1998-03-26
/ PRIOR APPLICATION NUMBER: 60/079728
/ PRIOR FILING DATE: 1998-03-27
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 246
/ SEQ ID NO 80
/ LENGTH: 667
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-230-338-80

Query Match          98.8%; Score 1048; DB 9; Length 667;
Best Local Similarity 99.0%; Pred. No. 6.5e-94;
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLIL 61
Db 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLIL 62
Qy 62 VLLILAMLVRRRLQMPDCVGRGRLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 121
Db 63 VLLILAMLVRRRLQMPDCVGRGRLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 122
Qy 122 ALPFTLLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSLTSLWA 181
Db 123 ALPFTLLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSLTSLWA 182
Qy 182 HLGVOVWQRAECPOVPKI 199
Db 183 HLGVOVWQRAECPOVPKI 200

RESULT 8
US-09-759-056-5
/ Sequence 5. Application US/09759056
/ Patent No. US20020156252A1
/ GENERAL INFORMATION:
/ APPLICANT: Pennica, Diane
/ APPLICANT: Smith, Victoria
/ APPLICANT: Wood, William I.
/ TITLE OF INVENTION: GENENT.2827A2
/ FILE REFERENCE: GENENT.2827A2
/ CURRENT APPLICATION NUMBER: US/09/759,056
/ PRIOR APPLICATION NUMBER: 2001-01-11
/ PRIOR FILING DATE: 2001-04-14
/ PRIOR APPLICATION NUMBER: 60/175849
/ PRIOR FILING DATE: 2000-01-13
/ PRIOR APPLICATION NUMBER: 60/228914
/ PRIOR FILING DATE: 2000-06-29
/ NUMBER OF SEQ ID NOS: 5
/ SOFTWARE: fastseq for Windows Version 4.0
/ SEQ ID NO 5
/ LENGTH: 658
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-759-056-5

Query Match          93.0%; Score 986.5; DB 9; Length 658;
Best Local Similarity 94.4%; Pred. No. 6.1e-88;
Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;
```

```
Qy 2 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLIL 61
Db 3 SOPAGNOTSPGATEDYSGSWYIDEPQGXELQPEGEVPSCHTSIPPGLYHACLASLIL 62
Qy 62 VLLILAMLVRRRLQMPDCVGRGRLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 121
Db 63 VLLILAMLVRRRLQMPDCVGRGRLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 113
Qy 122 ALPFTLLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSLTSLWA 181
Db 114 ALPFTLLASAPSDGKTEAPRGAMKILGLFHYAALYYPLAACATAGHTAAHLGSLTSLWA 173
```

QY 182 HLGQVWQRAECPOVPKI 199  
 |||||  
 Db 174 HLGQVWQRAECPOVPKI 191

## RESULT 9

US-09-901-812-5  
 ; Sequence 5, Application US/09901812  
 ; Patent No. US20020173461A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pennica, Diane  
 ; APPLICANT: Polakis, Paul  
 ; APPLICANT: Szeto, Wayne  
 ; TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO  
 ; FILE REFERENCE: GENENT.083A  
 ; CURRENT APPLICATION NUMBER: US/09/901.812  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/228,914  
 ; PRIOR FILING DATE: 2000-08-29  
 ; PRIOR APPLICATION NUMBER: 09/759,056  
 ; PRIOR FILING DATE: 2001-01-11  
 ; PRIOR APPLICATION NUMBER: 60/175,849  
 ; PRIOR FILING DATE: 2000-01-13  
 ; PRIOR APPLICATION NUMBER: 60/197,089  
 ; PRIOR FILING DATE: 2000-04-14  
 ; NUMBER OF SEQ ID NOS: 28  
 ; SOFTWARE: FaSTSeq for Windows Version 4.0  
 ; SEQ ID NO 5  
 ; LENGTH: 658  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-901-812-5

Query Match 93.0%; Score 986.5; DB 9; Length 658;  
 Best Local Similarity 94.4%; Pred. No. 6,1e-88;  
 Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

QY 2 SOPAGNTSGATGDEYSYGSWYIDEPQGXELQPEGEVPSCHTSLPPGLYHACLASLSTL 61  
 Db 3 SOPAGNTSGATGDEYSYGSWYIDEPQGXELQPEGEVPSCHTSLPPGLYHACLASLSTL 62  
 QY 62 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPAAVPAVFMVLLSSCLLPDED 121  
 Db 63 VLLLLAMLVRRQLMPDCVGRGRLPSVDFLAGDRPAAVPAVFMVLLSSCLLPDED 113  
 QY 122 ALPELTASPSQDGKTEAPRGAWKILGFHYAALYYPLAACATAGHTAAHLGSLTSLWA 181  
 Db 114 ALPELTASPSQDGKTEAPRGAWKILGFHYAALYYPLAACATAGHTAAHLGSLTSLWA 173  
 QY 182 HLGQVWQRAECPOVPKI 199  
 |||||  
 Db 174 HLGQVWQRAECPOVPKI 191

RESULT 10  
 US-09-816-653a-7  
 ; Sequence 7, Application US/0981653A  
 ; Publication No. US2003021788A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rastelli, Luca  
 ; APPLICANT: Pennica, Diane  
 ; TITLE OF INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME  
 ; FILE REFERENCE: 10716/57  
 ; CURRENT APPLICATION NUMBER: US/09/816,653A  
 ; PRIOR FILING DATE: 2001-03-23  
 ; PRIOR APPLICATION NUMBER: 60/191,532  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 7  
 ; LENGTH: 670

TYPE: PRT  
 ; ORGANISM: Mus musculus retinotic acid-responsive protein (Stra6) mRNA, complete cds  
 ; US-09-816-653a-7

Query Match 65.7%; Score 697; DB 9; Length 670;  
 Best Local Similarity 68.6%; Pred. No. 8.5e-60;  
 Matches 133; Conservative 18; Mismatches 41; Indels 2; Gaps 1;

QY 6 GNOTSPGATGDEYSYGSWYIDEPQGXELQPEGEVPSCHTSLPPGLYHACLASLSTLVL 65  
 Db 9 GSQSSGVTDYDS--SWIEEPLDAEVEQEGVIFLCQITAPALHACLASLSTLVL 66  
 QY 66 LAMLVRRQLMPDCVGRGRLPSVDFLAGDRPAAVPAVFMVLLSSCLLPDEDALPF 125  
 Db 67 LAMLVRRQLMPDCVGRGRLPSVDFLAGDRPAAVPAVFMVLLSSCLLPDENPLPF 126  
 QY 126 LTLASPSQDGKTEAPRGAWKILGFHYAALYYPLAACATAGHTAAHLGSLTSLWAHLGV 185  
 Db 127 LMTLASPSQDGKTEAPRGAWKILGFHYAALYYPLAACATAGHTAAHLGSLTSLWAHLGV 186  
 QY 186 QWQRAECPOVPKI 199  
 |||||  
 Db 187 QWQRAECPOVPKI 200

## RESULT 11

US-09-925-297-738  
 ; Sequence 738, Application US/09925297  
 ; Patent No. US20020081659A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: PA105  
 ; CURRENT APPLICATION NUMBER: US/09/925,297  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05989  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 928  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 738  
 ; LENGTH: 624  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (2)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (188)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (192)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-09-925-297-738

Query Match 7.9%; Score 84; DB 10; Length 624;  
 Best Local Similarity 22.7%; Pred. No. 3;  
 Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

QY 7 NQTSFGATGDEYSYGSWYIDEPQ---GXELQPEGEVPSCHTSLPPGLYHACLASLSTL 61  
 Db 112 NNSPFGATGTFPPAPWMPGPPPTNFSSLELPRGQOPVAKAGSPFALLIGLVAILLL 171  
 QY 62 VLLLLAMLV-----RRQLMPD-----CVGRGRLPSVDFL----- 93  
 Db 172 LLLIALLMLRMLHMRXLSKERRVLEELTVHLSPGDTLLINRRPGPREPPYQEPRP 231  
 QY 94 AGDRPAAVPAVFMVLLSSCLLPDEDALPFLTLASPSQDGKTEAPRGAW 145  
 Db 232 RGNPPHSAPC---VNGSALLSNPARYRLILATYARPGRG---PPTAW 276

```

RESULT 12
US-09-223-490-4
; Sequence 4, Application US/09223490
; Patent No. US20020147325A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Scadden, David T.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Barton, Will F.
; TITLE OF INVENTION: Protein Tyrosine Kinases
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: pafin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/223,490
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170,558
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haasek, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-223-490-4

Query Match          7.8%; Score 83; DB 10; Length 913;
Best Local Similarity 22.7%; Pred. No. 5.9;
Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

Oy 7 NOTSPGATEVDSYSGSWYIDBPQ---GGXELQEGEVPSCH-TATPRVLYNACIASLIL 61
Db 370 NNSPPLAGCTFPAPWPPPPPTNSSLLEPRGQGVAKPESPTALIGCLVATLIL 429
Oy 62 VLLILMLV-----RRQLMPD-----CVGRPGLSPPVDFL----- 93
Db 430 LLLIALLMLRLMRLLSKAERRVLEELTVHLSPGDTLLINRRPGEPPPYQEPFR 489
Oy 94 AGDRPRVPAVFAFMVLLSLCLLPDEDAFLTLASPSQDGKTEAPRGAM 145
Db 490 RGNPRHAPC-----VPRGSLLLSNPAVLLLATYARPPRPG---PPTPAW 534

RESULT 13
US-09-808-880-4
; Sequence 4, Application US/0980880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Belkash, Mary C.
; APPLICANT: Shah, Sanjay Kriemhakan
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li

```

```

; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 3062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3519
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant
US-09-808-880-4

Query Match          7.4%; Score 78.5; DB 9; Length 3519;
Best Local Similarity 23.5%; Pred. No. 82;
Matches 54; Conservative 25; Mismatches 70; Indels 81; Gaps 12;

Oy 11 PGATEDSYSGSWYIDBPQGXELQEGEVPSCHTSPPGLYHACLASLIVLLILMLV 70
Db 2642 PALDEVLAGGWLFFVPRG-----LADGVAHVAAVTA 2676
Oy 71 R-----RRQLMPDCVGRPGLPSPVDFLAGRPR---AVPA--AVFWTL 109
Db 2677 RGEVSVVELDPRPRKRAVAAVAGR-GVSGVVSFLSMDRKHSEHPVPFAGLASLVL 2735
Oy 110 LSSLC-----LLPDEDAFLTLASPSQDGKTEAPRGA--W---KILGLFH-- 152
Db 2736 AQAIVDLGRVGECPRLMLVTRDA-----VVAGPSDAGAVIDPVQAGVWGFGVGLGHEPD 2790
Oy 153 --YAALVYPLAA---CATAGHTAHLGLSTLSMAHL-----GVQVWOR 190
Db 2791 LMGSLIDLPVEAEPESGTDHTYADLLATVVASAGFEDQVAVAGSGVWR 2840

RESULT 14
US-09-738-626-6358
; Sequence 6358, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 6358
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum

```

US-09-738-626-6358

Query Match 7.0%; Score 74.5; DB 9; Length 208;

Best Local Similarity 26.5%; Pred. No. 6.7; Mismatches 25; Indels 43; Gaps 11;

Matches 50; Conservative 71; Indels 43; Gaps 11;

QY 21 SWYID-----EPQGXELQPEG---EVPSCHTSI---PGLYHACIASISILVL 63

DB 26 AMFVDTSLALDLINTSRAPCEVELSSQKASINLISLCSQSLPPRRLTSATTTSMSPLLT 85

QY 64 L--LIAMLYRRRLQMPDCVGRPGLPSPVDFLAGDRPRAVPAAVFVLLSLCLLPDED 121

DB 86 LKVSASKIVRTSCAMPLLFTRLIRSPVT-VASSLETPNLSASMKMAPS--CLVWKSLP 143

QY 122 ALPFLTLASAPSQDGTEAPRGAWKIT--GLFHYAALYPLAACATAGHTAAHLIGSTL- 178

DB 144 QIPPAT-----GLWKMKQDLCNQALI---SNSCAT--GSPALNLIAGTRIP 184

QY 179 SWAHLGVQY 187

DB 185 AWSHQLQY 193

## RESULT 15

US-09-925-301-1349

Sequence 1349; Application US/09925301

Patent No. US2002052308A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA106

CURRENT APPLICATION NUMBER: US/09/925,301

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05882

PRIOR FILING DATE: 2000-03-06

PRIOR APPLICATION NUMBER: 60/124,270

NUMBER OF SEQ ID NOS: 1694

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1349

LENGTH: 326

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (137)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (142)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-301-1349

Query Match 7.0%; Score 74; DB 10; Length 326;

Best Local Similarity 28.0%; Pred. No. 13; Mismatches 58; Conservative 16; Indels 79; Gaps 12;

QY 38 EVPSCHTSI-----PGUYHACIASISILVLLIAMLVRRQLMPDCVGRGRLP 87

DB 39 DVACPGSLCALKRARCPRA-HACGFCLOPPEDOGICVPRMRPPGGRPPRL 97

QY 88 SPVDFLA-----GDRPR-AVPAV-FVLLISLCLLPDEDALPILT-- 127

DB 98 DEIDFLAQLARKEGSHTPPLPKORRLPEPATIGFSAXGCGXLGI.PSTPGTPPTPH 157

QY 128 --LASAPSDG--KTAPRGAMKILGLFHYAALYPLAACATAGHTAAHLIGSTLSMAH 182

DB 158 TSLGSPVSSDPVMSPLERIG-----GQDGLALVILAF-C-VAG--AAALSVASLCMR 209

QY 183 L--GVQVWQRAE-----CPQVPI 199

DB 210 LQREIRILQKADYATAKGSPAPRI 236

Search completed: March 13, 2003, 11:37:57  
Job time : 12.355 secs





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:26:32 ; Search time 20.8874 Seconds  
(without alignments)  
1963.067 Million cell updates/sec

Title: US-09-816-653A-2  
Perfect score: 1061  
Sequence: 1 MSQACGQTSPGATEDSYG.....MAHGVQVWQRAECRPYK1 199

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:.\*  
2: SP\_BACTERIA:.\*  
3: SP\_FUNGI:.\*  
4: SP\_HUMAN:.\*  
5: SP\_INVERTEBRATE:.\*  
6: SP\_MAMMAL:.\*  
7: SP\_MHC:.\*  
8: SP\_ORGANELLE:.\*  
9: SP\_PHAGE:.\*  
10: SP\_PLANT:.\*  
11: SP\_PROTOZOA:.\*  
12: SP\_VIRUS:.\*  
13: SP\_VERTEBRATE:.\*  
14: SP\_UNCLASSIFIED:.\*  
15: SP\_VIRUS:.\*  
16: SP\_BACTERIAP:.\*  
17: SP\_ARCHAEP:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1048	98.8	667	4 Q9BX79	Q9BX79 homo sapien
2	1048	98.8	667	4 Q9BX79	Q9BX79 homo sapien
3	986.5	93.0	658	4 Q9BX78	Q9BX78 homo sapien
4	697	65.7	670	11 Q70491	Q70491 mus musculu
5	90	8.5	263	5 Q8SUI0	Q8SUI0 encephalit
6	89.5	8.4	750	12 Q9YQ39	Q9YQ39 pseudorabie
7	87.5	8.2	336	10 Q94LD3	Q94LD3 oryza sativ
8	87.5	8.2	413	16 Q8XRV5	Q8XRV5 ralsstonia s
9	86.5	8.2	757	16 Q9AA25	Q9AA25 caulobacter
10	85.5	8.1	569	16 Q8Y1X8	Q8Y1X8 ralsstonia s
11	84.5	8.0	730	12 Q930X7	Q930X7 corynebacte
12	84.5	8.0	730	12 Q930X7	Q930X7 streptomyce
13	83	7.8	1172	16 Q9ZBY5	Q9ZBY5 lassaia sp.
14	82	7.7	208	8 Q36437	Q36437 homo sapien
15	82	7.7	2641	4 Q9BXD4	Q9BXD4 streptomyce
16	81.5	7.7	9507	2 Q9EWAL	Q9EWAL streptomyce

17	81	7.6	278	11 Q91YD1	Q91YD1 mus musculu
18	81	7.6	458	2 Q9F2A0	Q9F2A0 chromatium
19	80.5	7.6	120	17 Q8ZUY6	Q8ZUY6 pyrobaculum
20	80.5	7.6	241	2 Q8VPR4	Q8VPR4 micrococcus
21	80.5	7.6	385	16 Q91AM7	Q91AM7 pseudomonas
22	79.5	7.5	151	11 Q9CYX5	Q9CYX5 mus musculu
23	79.5	7.5	249	2 Q57382	Q57382 escherichia
24	79.5	7.5	664	16 Q9F2V5	Q9F2V5 streptomyce
25	79	7.4	433	12 Q9BS90	Q9BS90 prospect hi
26	79	7.4	579	10 Q9FMC8	Q9FMC8 oryza sativ
27	79	7.4	1312	3 Q8WZV2	Q8WZV2 neurospora
28	78.5	7.4	447	4 Q9Y220	Q9Y220 homo sapien
29	78.5	7.4	477	16 Q91660	Q91660 pseudomonas
30	78.5	7.4	601	16 Q8Y3B6	Q8Y3B6 ralsstonia s
31	77.5	7.3	156	10 Q8R258	Q8R258 oryza sativ
32	77.5	7.3	439	16 Q9F2N1	Q9F2N1 streptomyce
33	77.5	7.3	476	16 Q8XZ18	Q8XZ18 ralsstonia s
34	77.5	7.3	1311	6 Q9BDH6	Q9BDH6 equus caball
35	77	7.3	521	16 Q8YJ03	Q8YJ03 bruceella me
36	76.5	7.2	280	2 Q8RTX3	Q8RTX3 xanthomonas
37	76.5	7.2	311	3 Q8WZM2	Q8WZM2 trichoderma
38	76.5	7.2	406	10 Q9FRB9	Q9FRB9 oryza sativ
39	76	7.2	128	4 Q96DK3	Q96DK3 homo sapien
40	76	7.2	203	8 Q33884	Q33884 arabidopsis
41	76	7.2	394	16 Q53924	Q53924 streptomyce
42	76	7.2	539	10 Q8S7M1	Q8S7M1 oryza sativ
43	76	7.2	666	16 Q91I43	Q91I43 pseudomonas
44	76	7.2	709	16 Q9RYK7	Q9RYK7 deinococcus
45	75.5	7.1	360	11 Q8R4S8	Q8R4S8 rectus novy

## ALIGNMENTS

### RESULT 1

Q9BX79 PRELIMINARY; PRT; 667 AA.  
ID Q9BX79 Q9BX79; Q9H9U8;  
AC Q9BX79; Q9H9U8; 17, Created)  
DT 01-JUN-2001 (TRENBLER1, 17, Last sequence update)  
DT 01-JUN-2002 (TRENBLER1, 21, Last annotation update)  
DE STRA6 isoform 1.  
OS Homo sapiens (Human).  
OC Mammalia; Euteleostomi; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteleostomi; Chordata; Vertebrata; Euteleostomi;  
OC NCB1\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Saito M., Jiang W., Tice D.A., Rubinfield B., Hollingshead P.G.,  
RA Feng S.B., Dugger D.L., Pham T., Yansura D.G., Wong T.A.,  
RA Grimaldi J.C., Corpuz R.T., Singh J.S., Frantz G.D., Devaux B.,  
RA Crowley C.W., Schwall R.H., Eberhard D.A., Rastelli L., Polakis P.,  
RA Pennica D.;  
RT "Overexpression of the retinoic acid-responsive gene Stra6 in human  
RT cancers and its synergistic induction by Wnt-1 and retinoic acid";  
RT Cancer Res. 0:0-0(2001).  
RL [2]  
RN SEQUENCE OF 108-667 FROM N.A.  
RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Iehi S., Kawai Y., Saito K.,  
RA Yamamoto J., Watanabe A., Nakamura Y., Nagahara K., Masuho Y.,  
RA Niimura K., Iwayanagi T.;  
RT "NED0 human cDNA sequencing project";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF52728; AK30289.1; -;  
DR EMBL; AK022603; BAB1412.1; -;  
DR VARIANT 527 M -> I.  
SQ SEQUENCE 667 AA; 73502 MW; D20840A46998BA2E CRC64;  
Query Match 98.8%; Score 1048; DB 4; Length 667;

Best Local Similarity 99.0%; Pred. No. 3.3e-95;  
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSPGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSTL 61  
DB 3 SOPAGNOTSPGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSTL 62  
QY 62 VLLLLAMLVRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSSLCLLLPDED 121  
DB 63 VLLLLAMLVRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSSLCLLLPDED 122  
QY 122 ALPFLTLASAPSDGKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLMA 181  
DB 123 ALPFLTLASAPSDGKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLMA 182  
QY 182 HLGVOVMORAECPQVPKI 199  
DB 183 HLGVOVMORAECPQVPKI 200

RESULT 2  
Q8TB21 PRELIMINARY; PRT; 667 AA.  
ID Q8TB21  
AC Q8TB21  
DT 01-JUN-2002 (Tremblrel. 21, Created)  
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Hypothetical 73.5 kDa protein.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Strauberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC025256; AAH25256.1; -  
KW Hypothetical protein  
SQ SEQUENCE 667 AA; 73532 MW; 120685F4635CB0B5 CRC64;

Query Match 98.8%; Score 1048; DB 4; Length 667;  
Best Local Similarity 99.0%; Pred. No. 3.3e-95;  
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SOPAGNOTSPGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSTL 61  
DB 3 SOPAGNOTSPGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSTL 62  
QY 62 VLLLLAMLVRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSSLCLLLPDED 121  
DB 63 VLLLLAMLVRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSSLCLLLPDED 122  
QY 122 ALPFLTLASAPSDGKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLMA 181  
DB 123 ALPFLTLASAPSDGKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLMA 182  
QY 182 HLGVOVMORAECPQVPKI 199  
DB 183 HLGVOVMORAECPQVPKI 200

RESULT 3  
Q9BX78 PRELIMINARY; PRT; 658 AA.  
ID Q9BX78  
AC Q9BX78  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE STRA6 Isoform 2.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Szeto W., Jiang W., Tice D.A., Rubinfeld B., Hollingshead P.G.,  
RA Fong S.E., Dugger D.L., Pham T., Yansura D.G., Wong T.A.,  
RA Grimaldi J.C., Corpuz R.T., Singh J.S., Frenet G.D., Devaux B.,  
RA Crowley C.W., Schwall R.H., Eberhard D.A., Rastelli L., Polakis P.,  
RA Pennica D.;  
RT "Overexpression of the retinoic acid-responsive gene Stra6 in human  
RT cancers and its synergistic induction by Mnt-1 and retinoic acid";  
RL Cancer Res. 0:0-0(2001)  
DR EMBL; AF352729; AAK30290.1; -  
FT VARIANT 518 518  
SQ SEQUENCE 658 AA; 72582 MW; 5405281727647D CRC64;

Query Match 93.0%; Score 986.5; DB 4; Length 658;  
Best Local Similarity 94.4%; Pred. No. 3.9e-89;  
Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

QY 2 SOPAGNOTSPGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSTL 61  
DB 3 SOPAGNOTSPGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSTL 62  
QY 62 VLLLLAMLVRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSSLCLLLPDED 121  
DB 63 VLLLLAMLVRRQLMPDCVGRPGLPSPVDFLAGDRPAVPAVFWVLLSSLCLLLPDED 113  
QY 122 ALPFLTLASAPSDGKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLMA 181  
DB 114 ALPFLTLASAPSDGKTEAPRGAMKILGLFYAALYYPPLAACATAGHTAAHLGSLTSLMA 173  
QY 182 HLGVOVMORAECPQVPKI 199  
DB 174 HLGVOVMORAECPQVPKI 191

RESULT 4  
O70491 PRELIMINARY; PRT; 670 AA.  
ID O70491  
AC O70491  
DT 01-AUG-1998 (Tremblrel. 07, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE Retinoic acid-responsive protein.  
GN STRA6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9537753; PubMed=7649373;  
RA Bouillet P., Oulad-Abdelhann M., Vicaire S., Garnier J.M.,  
RA Schaubert B., Dolle P., Chambon P.;  
RT "Efficient cloning of cDNAs of retinoic acid-responsive genes in P19  
RT embryonal carcinoma cells and characterization of a novel mouse gene,  
RT Stral (mouse LERK-2/Ep192).";  
RL Dev. Biol. 170:420-433(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97346723; PubMed=9203140;  
RA Bouillet P., Sapin V., Chazaud C., Messadeg N., Decimo D., Dolle P.,  
RA Chambon P.;  
RT "Developmental expression pattern of Stra6, a retinoic acid-responsive  
RT gene encoding a new type of membrane protein.";  
RL Mech. Dev. 63:173-186(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Bouillet P.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF062476; AAC16016.1; -  
MGD; MGI:107742; Stra6.  
SQ SEQUENCE 670 AA; 73774 MW; 6ACE69F3BE16A4F7 CRC64;



```

Query Match      8.2%; Score 87.5; DB 10; Length 336;
Best Local Similarity 34.7%; Pred. No. 1.3;
Matches 35; Conservative 13; Mismatches 42; Indels 11; Gaps 5;

QY 43 HTSIPGLYHACLASLILVLLMLV---RRQLPDVCGRGPG-PSVPDLADDR 97
DB 211 HRRISGVFHAVLVSVQLLAVPRRLLAGRGRRRRRRPPVPGRGVRPSVKPSSVVR 270

QY 98 PRAV-----PAVFWLLSSCLLIPDDAL-PLTLASAPS 133
DB 271 VRQVCRSPVVVF-VLASASSLVPAASRLPRIAAEVVS 310

RESULT 8
Q8XRV5 PRELIMINARY; PRT; 413 AA.
AC Q8XRV5:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Putative MFS transporter transmembrane protein.
GN RSP0726 OR R801724.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_Taxid=305;
RA SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Gaspin C., Lavie M., Moisan A., Claudel-Renard C., Cunne S., Demange N.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646080; CAD17877.1;
KW plasmid; Complete proteome.
SQ SEQUENCE 413 AA; 43496 MW; F4C0B6453C037CC64;

Query Match      8.2%; Score 87.5; DB 16; Length 413;
Best Local Similarity 24.1%; Pred. No. 1.6;
Matches 47; Conservative 28; Mismatches 75; Indels 45; Gaps 8;

QY 1 MSOPAGNQTSPGATEDSYSGSWYIDPOGXLQPEGEVPSCHTSIPPGLYHA---CLAS 57
DB 134 LTESNGREAAAGVNRALTYASKLIPVVGSL-----IFIDYVALTCCSANV 181

QY 58 LSTIVLLMLVRRRLQWPDVCGRGPG-PSVPDLADDR-----PRAVPA-----A 104
DB 182 LALSAILLRSSIR-----CQSSRQNPAGKDDPFGIRKTTLSFYGGSSEBALDAVILTA 234

QY 105 VFWLLSSCLLIP-DEDALPFLTLASAPSDGKTEAPRGAMKILGFHYAALVPLAAC 163
DB 235 LFNFWLAPLVIIIPMIVSKLPPATSLST---GLSEASLGAAGVCG---ALLFRLRT 285

QY 164 ATAGHTAHLGLSTL 178
DB 286 GSPKHSLLVAGASL 300

RESULT 9
Q9AAZ5 PRELIMINARY; PRT; 757 AA.
AC Q9AAZ5:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Beta-N-acetylhexosaminidase, putative.

```

```

GN CC0447.
OC Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_Taxid=155892;
RA SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Pococka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwalt M.L., Haff D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Yamalhevan J., Emolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005718; AAK22434.1;
DR HSSP; P06865; 10BC.
DR TIGR; CC0447;
DR InterPro; IPR001540; GH_20.
DR Pfam; PF00728; Glyco_hydro_20.1.
DR Pfam; PF02838; Glyco_hydro_20b.1.
DR PRINTS; PR00738; GLHYDRLASE20.
KW Complete proteome.
SQ SEQUENCE 757 AA; 82171 MW; 7E24C4E7EB173803 CRC64;

```

```

Query Match      8.2%; Score 86.5; DB 16; Length 757;
Best Local Similarity 30.6%; Pred. No. 4;
Matches 49; Conservative 16; Mismatches 60; Indels 35; Gaps 9;

```

```

QY 62 VLLMLVLR-RQLMPDVCGRGPG-PSVPDLADDR-----RAVPA 103
DB 65 VALQLDLVLRGRGRPKVVEGPPADAIV--LTREBPAGEAYKLDINKGATIAAKRA 122

QY 104 AVEFWLLSSCLLIPDEDALPFLTLASAPSDGKTEAPRGAMKILGFHYAALVY-----P 159
DB 123 GLPYGAMSLQWQATPDPAKGPVALLAISIE-----DAKPRVNR--GLWDSARHQSQSD 175

QY 160 LACATAGHTAHLGLSTLSNAHLGVQVQRAECPQPKI 199
DB 176 LKAVIDA--MAAHKL-NTFHMLVDDQW-RLETKYPKL 211

RESULT 10
Q8YIX8 PRELIMINARY; PRT; 569 AA.
AC Q8YIX8:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical transmembrane protein RSC0559.
GN RSC0559 OR RS04906.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_Taxid=305;
RA SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Gaspin C., Choise N., Claudel-Renard C., Cunne S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646060; CAD14089.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 569 AA; 62435 MW; C1AE4805E1A843EE CRC64;

```

Query Match	8.1%	Score 85.5;	DB 16;	Length 569;
Best Local Similarity	24.1%	Pred. No. 3.6;		
Matches	49;	Conservative	21;	Mismatches 60;
				Indels 73;
				Gaps 9;
Qy	49	GLHACIASLSTIVLLTMLVRRRQQLDVRGRPLP-----SPVDFLAG-----	95	
Db	56	GLRPVLIAPLALGVILQVLYV-----ATPLNPFTFAAIYFLLGTVAVCGLG	104	
Qy	96	-----DRPAVPAVAFVNTLSL-----CLLPDEDALEPFLTAAPQDGTSPAGAMK	146	
Db	105	RCRDPEGVLSIAVNIIGLLVTAIECHLPRVSGLIINMISLP-----TGAGRRMG	159	
Qy	147	TLGFTFAIYLL--YPLAACATAGT-----AAHLG-----STLSNAHL-----	183	
Db	160	MLNPQHVATYLLAGLAACLPLOSTTRRRVAPLAIATALLGLPALTVSRMSNHLVLYG	219	
Qy	184	-----GVQVQRAECP	194	
Db	220	GVAGLMSAEERGAARRIRAGVP	242	

```

RESULT 11
ID 039505 PRELIMINARY; PRT; 382 AA.
AC 039505;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE Glycoprotein I.
GN US7.
OS Bovine herpesvirus 1.
OC dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirinae.
OX NCBI_Taxid=103320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Goltz M., Buhk H.J., Broll H., Lewin M., Mankertz A., Boerner B.,
RA Borchers K., Weigelt W.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 372-382 FROM N.A.
RC STRAIN=JURA;
RA Schwyzer M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwyzer M.;
RA Lowery D.E., Simard C., Bello L.J., Thiry E., Vlcek C.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Goltz M., Buhk H.J., Schwyzer M.;
RT "Nucleotide sequence of the HindIII O and K fragments located in the
RT US region of the bovine herpesvirus 1 genome."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL, AJ004801; CAA046.1; -.
DR EMBL, Z98199; CAB10890.1; -.
DR InterPro; IPR002874; Herpes_gI.
DR Pfam; PF01688; Herpes_gI; 1.
SQ SEQUENCE 382 AA; 39596 MW; 0F06C95CD2581EC2 CRC64;

Query Match 8.0%; Score 84.5; DB 12; Length 382;
Best local similarity 27.6%; Pred. No. 2, 9;
Matches 37; Conservative 12; Mismatches 38; Indels 47; Gaps 6

4 PAGNGTSPGATDYSYGSWYIDPEQGGKGLDPGSEVPSCHTSIPGLVNAICLAS----LS 59
258 PAVLTNPSPAASPSRAFSN-----AAPMAAAGPAGDTPA-----RRRQLASLIVPLC 304

```

[illegible]

RESULT 12			
ID	0930X7	PRELIMINARY;	PRT; 730 AA.
AC	Q930X7;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Hypothetical 77.3 kDa protein.		
OS	Corynebacterium glutamicum (Brevibacterium flavum).		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Corynebacterineae; Corynebacteriaceae;		
OC	Corynebacterium.		
OX	NCBI_TaxID=1718;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MD233;		
RA	Kubasak P., Kobayashi M., Sandmann G.;		
RT	"Expression and functional analysis of a gene cluster involved in the		
RT	synthesis of decaprenoxanthin reveals the mechanisms for C50		
RT	carotenoid formation."		
RL	Bur. J. Biochem. 286:3702-3708(2001).		
DR	EMBL; AF159510; AA64297.1; ..		
DR	InterPro; IPR001092; HLM_Basic.		
DR	InterPro; IPR000731; HMGCR/patch_5TM.		
DR	InterPro; IPR004869; HTH_LysR.		
DR	InterPro; IPR004869; MPMPL.		
DR	Pfam; PF03116; MPMPL; 2.		
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.		
DR	PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.		
DR	PROSITE; PSS0156; SSD_1.		
DR	Hypothetical protein.		
SQ	SEQUENCE 730 AA; 77290 MW; 18A9CAE088418633 CRC64;		

```

Query March 8.0%; Score 84.5; DB 2; Length 730;
Best Local Similarity 29.4%; Pred. No. 6;
Matches 30; Conservative 18; Mismatches 41; Indels 13; Gaps 3

Oy 49 GYHACIASLILVLLI-----LAMLVRRRLMPDCVGRGSLP-----SPVDFLAGDRP 98
| : : : : : | : : : : : |
Db 319 GIIFMMSALTLLEALLFVFGRAVFWPKRPKPEVEVGENGIDPAGSIWKAHLVEQHP 378
| : : : : : | : : : : : |

Oy 99 RAVPAVPMVLSSLC---LLPEDNALPFLTLASAPQDPK 137
| : : : : : | : : : : : |
Db 379 RAIWSTLIVLLLGAAVFPFLTKADGVSGSDLVLGSSSEARDQ 420
| : : : : : | : : : : : |

RESULT 13
O9ZBY5 PRELIMINARY; PRT; 1172 AA.
ID O9ZBY5;
AC O9ZBY5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative ABC transporter.
DB CYDCD OR SC03947 OR SC078.14.
OS Streptomyces coelicolor.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_Taxid:1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

```

```

RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RX Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered contigs and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challin G.L.,
RA Thomson N.R., James K.D., Harris D.B., Quail M.A., Kleiser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietorrek A., Woodward J., Barrett B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
RN [15]
RP STIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
RC EMBL; AL034355; CAI22219.1; -.
RA HSSP; P13569; INBD.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR001140; ABCtransporterTM.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR000524; HTH_GntR.
DR Pfam; PF00664; ABC_membrane; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; UNKNOWN_1.
DR ATP-binding; Transport.
SQ SEQUENCE 1172 AA; 122768 MW; 57101CD9326D704D CRC64;

Query Match
Best Local Similarity 7.8%; Score 83; DB 16; Length 1172;
Matches 45; Conservative 12; Mismatches 68; Indels 18; Gaps 5;

QY 55 LASISIVLVLLMLVRRQLWPCVGRPGPLSPVDPLACDRPRAVPAVFWLLSSLC 114
DB 674 LAABEAVLGLPLAVQYRQR-VRSARVYEVLDAPREPEVRQAAPTFPVVEGIA 932
QY 115 LLLP--DEDALEPLTL-----ASAPSDGKTEADRGWKILGFHYAALYPLA-- 161
DB 933 ARHGGORDALAGLDITLGGRRVAVVVGSGSGKTTL---AQVLRPLDSAGSYLAGY 989
QY 162 -ACATGHTAAHLTGSTLSMAHL 183
DB 990 DAYALAGDVRLVGLCAODAH 1012

RESULT 14
Q36437 PRELIMINARY; PRT; 208 AA.
AC Q36437;
DT 01-JUN-1996 (TREMBlrel. 01, Created)
DT 01-JUN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cytochrome oxidase III protein.
OS Lasea sp.
OG Mitochondrion.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Heteroconchia; Veneroidea;
OC Galeommatoidae; Laseidae; Lasea.

```

```

OX NCBI_TaxID=32592;
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=KRG;
RA O'Folghid D., Smith M.J.;
RL "Evolution of asexuality in the cosmopolitan marine clam Lasea.";
RT Evolution 49:140-150(1995).
DR EMBL; X78301; CAAS5106.1; -.
DR InterPro: IPR000298; Cytochrome_oxidase_III.
DR Pfam; PF00510; COX3; 1.
DR ProDom; PD000382; Cytochrome_oxidase_III; 2.
DR PROSITE; PS50253; COX3; 1.
DR Mitochondrion.
SQ SEQUENCE 208 AA; 22785 MW; 05517676CB39576 CRC64;

Query Match
Best Local Similarity 7.7%; Score 82; DB 8; Length 208;
Matches 33; Conservative 24; Mismatches 58; Indels 22; Gaps 5;

QY 60 ILVLLMLVRRQLWPCVGRPGPLSPVDPLACDRPRAVPAVFWLLSSLC 116
DB 12 VLLEFLAIFLGLTIAWGVIVKEGYLCHSLVA---KGLRGMAFLISEVFF 67
QY 117 LPEDALPLTLASAPSDGKTEADRGWKILGFHYAALYPLAACATGHTAHL-IG 175
DB 68 VSRFMAVFFHLISGLS-----AEGVMPMGIFPINAQVPL-----NTAVLISG 113
QY 176 STSWAHLGVQVORAE 192
DB 114 VSTWMAVVAIREWSRAQ 130

RESULT 15
Q9BXD4 PRELIMINARY; PRT; 2641 AA.
AC Q9BXD4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE PDZ domain-containing protein AIPC.
GN AIPC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21182735; PubMed=11289102;
RA Chab H., Rubin M.A., Mucci N.R., Li L., Taylor J.M.G., Day M.L.,
RA Rhim J.S., Macoska J.A.;
RT "Activated in prostate cancer: A PDZ domain-containing protein highly
expressed in human primary prostate tumors.";
RL Cancer Res. 61:2390-2394(2001).
DR EMBL; AF338650; AAK07661.1; -.
DR HSSP; Q14005; 1116.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001476; PDZ.
DR Pfam; PF00595; PDZ; 4.
DR SMART; SM00228; PDZ; 6.
DR PROSITE; PS00107; PDZ; 6.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
SQ SEQUENCE 2641 AA; 280205 MW; CBBF84582975754 CRC64;

Query Match
Best Local Similarity 7.7%; Score 82; DB 4; Length 2641;
Matches 50; Conservative 19; Mismatches 64; Indels 70; Gaps 9;

QY 11 PCATDYSY-GSWYIDEGGEXLOPGVPSCHSIP-----PGLYH- 52
DB 679 PGSLDPVAVASDEEDHSGCSTSESGISPFSTSEPKPRANSLVTLGSHRASGLPHK 738
QY 53 ----ACLASISILVLLMLVRRQLW-----DCVGRPGPLSPVD 92

```

```

Db      739 QYTVARQASLPGSPQALNPLLRQKVCYDANDASDEEFDRGDCISLPGALPGPIRP 798
Qy      93  LAGDRPR-----AVPAVFWLJSSLCLLLPDEDALP--FLTJ 128
      | : | | |
Db      799 LSEDDPRRVYISSSKGMVDHNOERPRKTIWKAISAPLIGSSVDL--EESIPEGWVDA 855
      | : | | |
Qy      129 AS-APSGDGKTEAPRGA---WK 146
      | | | : | | |
Db      856 ASYAAALTDSEAPKSGSPGSMWK 878
      | | | : | | |

```

Search completed: March 13, 2003, 11:29:56  
 Job time : 25.8874 secs





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:26:57 ; Search time 6.45611 Seconds  
(without alignments)  
1278.447 Million cell updates/sec

Title: US-09-816-653a-2

Perfect score: 1061  
Sequence: 1 MSQPDAGNGTSPGATBEDYSYG.....MAHGVQWQRECPQVFKI 199

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	8.1	380	VGLI_HSVBS	008102 bovine herp
2	84.5	8.0	335	TRD2_STRCO	0924w streptomyce
3	84	7.9	913	DDRI_HUMAN	008345 h epichelila
4	83	7.8	281	CBSE_RAT	P56261 rattus norv
5	80	7.5	433	NCAP_PHV	P22047 prospect hi
6	80	7.5	925	NFC2_HUMAN	013469 homo sapien
7	79	7.4	910	DDRI_RAT	063474 rattus norv
8	78.5	7.4	532	GRB7_HUMAN	014451 homo sapien
9	78.5	7.4	577	CCBS_OENBE	004648 oenothera b
10	78.5	7.4	3519	OL56_STRAT	007017 streptomyce
11	78	7.4	281	CBSE_HUMAN	015744 homo sapien
12	78	7.4	911	PTB_PEA	004903 pisum sativ
13	78	7.4	919	DDRI_MOUSE	003146 mus musculu
14	77.5	7.3	579	CCBS_DAUCA	004647 daucus caro
15	76.5	7.2	874	PAG1_MOUSE	P97481 mus musculu
16	75.5	7.1	655	IDUA_CANFA	001634 canis fam11
17	74	7.0	867	POL_SRY2	P51517 simian retr
18	73	6.9	786	YVOI_MYCTU	050723 mycobacteri
19	72.5	6.8	138	ENV_TGFP	P15073 myc cell f
20	72	6.8	1385	NAC2_CHLNB	Q9188 chlamydomon
21	71.5	6.7	1844	POLR_TMYC	P28477 turnip yell
22	71	6.7	193	YAPT_PSEST	P28476 pseudomonas
23	71	6.7	325	NPD1_HUMAN	Q9nq45 homo sapien
24	70.5	6.6	430	NOR2_RAT	Q63516 rattus norv
25	70.5	6.6	888	NR43_RAT	P51179 rattus norv
26	70.5	6.6	888	YR43_RAT	P38016 escherichia
27	70	6.6	931	SMC_MOUSE	Q9wn33 mus musculu
28	70	6.6	1032	YAG0_SYNY3	P72637 synecocyst
29	70	6.6	1189	PTNB_MOUSE	Q62130 mus musculu
30	69.5	6.6	262	COX3_BRALA	P47403 brachyosco
31	69.5	6.6	329	IHA_CHICK	043301 gallus gall
32	69.5	6.6	425	OXIR_HUMAN	043613 homo sapien
33	69.5	6.6	465	FTSW_MYCLE	Q50186 mycobacteri

34	69.5	6.6	867	1	POL_MPMV	P07572 simian maso
35	69.5	6.6	1204	1	NO3_BOVIN	P29473 bos taurus
36	69.5	6.5	3412	1	POLG_TBREV	P07720 t genome po
37	68.5	6.5	194	1	FGF4_CHICK	P48804 gallus gall
38	68.5	6.5	273	1	DLX3_CHICK	P53770 notophthalm
39	68.5	6.5	452	1	MURD_PLEWA	Q91284 pleurodeles
40	68.5	6.5	452	1	CP7_ORYLA	P70085 oryzias lat
41	68.5	6.5	517	1	NIFJ_ECOLI	P52647 escherichia
42	68.5	6.5	1174	1	NIFJ_ECOLI	Q01098 mus musculu
43	68.5	6.5	1239	1	NME3_MOUSE	Q8xy07 ralestonia s
44	68	6.4	277	1	MUR1_PALSO	Q10554 mycobacteri
45	68	6.4	505	1	Y895_MYCTU	

## ALIGNMENTS

## RESULT 1

VGLI\_HSVBS STANDARD; PRT: 380 AA.

AC Q08102;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Glycoprotein I.  
GN GI.  
OS Bovine herpesvirus type 1.2 (strain ST).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirub.  
OX NCBI\_Taxid=45407;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94167875; PubMed=8122370;  
RA Leung-Trick P., Audonnet J.F., Riviere M.;  
RT "The complete DNA sequence and the genetic organization of the short  
unique region (NS) of the bovine herpesvirus type 1 (ST strain).";  
RL Virology 199;409-421(1994).  
-I- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS I, TO VZV GFLV,  
AND TO PRV GP63.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z33068; CAA80605.1; -  
DR InterPro: IPR002874; Herpes\_G1.  
DR Pfam: PF01688; Herpes\_G1; 1.  
KW Glycoprotein.

FT CARBOHYD 67 N-LINKED (GHCNAC... ) (POTENTIAL).  
SQ SEQUENCE 380 AA; 39910 MW; ABBELP89BA30D2BD CRC64;

Query Match 8.1%; Score 85.5; DB 1; Length 380;  
Best Local Similarity 27.6%; Pred. No. 1.2; 38; Indels 47; Gaps 6;  
Matches 37; Conservative 12; Mismatches

QY 4 PAGNONGSPGATDVGSGWYIDEPQGXELQPEGEVPSCHTSIPQLVYACLAS-----LS 59  
DB 256 PAPTPTSPAPASRAFSF-----AAPAAQAQPADDTA-----RFRROGLASTIVPLC 302  
QY 60 ILVLLILAM-----LVRRRL-----WPCVGRGPGSPV 90  
DB 303 VLVLLILACAAIVNCARLLPCSRVRVYKPTCAAGSGTCAGRPFGAAPSAPATV 362  
QY 91 DFLAGRPRAVPA 104  
DB 363 VAL-GSRPKAPPLA 375

RESULT 2

```

TRD2_STRCO STANDARD; PRT; 335 AA.
ID TRD2_STRCO STANDARD; PRT; 335 AA.
AC 0924W9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Anthranilate phosphoribosyltransferase 2 (EC 2.4.2.18).
GN TRD2 OR SCO3212 OR SCE8.05C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OK NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Charer K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleeser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Frazer A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleeser T., Laiké L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierhorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
CC -1- CATALYTIC ACTIVITY: Anthranilate + phosphoribosylpyrophosphate =
CC N-5-phosphoribosyl-anthranilate + diphosphate.
CC -1- PATHWAY: tryptophan biosynthesis; second step.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL035654; CAB38583.1;
CC InterPro: IPR000312; Glycose_transf_3.
CC Pfam: PF00591; Glycose_transf_3; 1.
CC DR Pfam: PF02865; Glycose_transf_3; 1.
CC DR Pfam: PF001864; Glycose_transf_3; 1.
CC DR TrEMBL: T1GR01245; trpd; 1.
CC KW Tryptophan biosynthesis; Transferase; Glycosyltransferase;
CC Complete proteome.
CC SEQUENCE 335 AA; 34536 MW; 2706194E400B2F0D CRC64;
SQ
Query Match 8.0%; Score 84.5; DB 1; Length 335;
Best Local Similarity 30.8%; Pred. No. 1.3;
Matches 36; Conservative 10; Mismatches 42; Indels 29; Gaps 5;
QY 81 KGRPGLPSPVPLAGDRPAVPAVFMVLLSLCLLPDEDAJPLTLASAPSGDKTEA 140
Db 247 KSRPG-----DIAGGD--FAVVAVALRVLAG-----GRGPARDVLLN 283
QY 141 PRAMWILGPFYVALYPLAACATAGHTAHLGSLTSMANHLGVQWQRAEPCQVP 197
Db 284 AAALRAVAGVAGTWDGLRLAASAVDGAAGGLD---RWALHS---WQRAALVEVP 334

```

```

DE (Tyrosine-protein kinase CAK) (Cell adhesion kinase) (Tyrosine kinase
DE DDR) (Discooidin receptor tyrosine kinase) (TRK E) (Protein-tyrosine
DE kinase Rtk 6) (CD167a antigen).
GN CN DDR1 OR EDDR1 OR CAK OR TRKE OR RTK6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RX MEDLINE=94134417; PubMed=8302582;
RA Perer J.L., Shen X., Finkenzel S., Sciortta L., Jenkins N.A.,
RA Gilbert D., Copeland N.G., Wong T.W.;
RT "Identification and chromosomal mapping of a receptor tyrosine kinase
RT with a putative phospholipid binding sequence in its ectodomain.";
RL Oncogene 9:211-219(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93296201; PubMed=8390675;
RA Johnson J.D., Edman J.C., Rutter W.J.;
RT "A receptor tyrosine kinase found in breast carcinoma cells has an
RT extracellular discooidin I-like domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5677-5681(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97131588; PubMed=8977099;
RA Sakuma S., Tada M., Sava H., Sawamura Y., Shinohre Y., Abe H.;
RT "Receptor protein tyrosine kinase DDR is up-regulated by p53
RT protein.";
RL FEBS Lett. 398:165-169(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=96389017; PubMed=8796349;
RA Playford M.P., Butler R.J., Wang X.C., Katso R.M., Cooke I.B.,
RA Ganesan T.S.;
RT "The genomic structure of discooidin receptor tyrosine kinase.";
RL Genome Res. 6:620-627(1996).
RN [5]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Lung;
RX MEDLINE=96204002; PubMed=8622863;
RA Perez J.L., Jing S.O., Wong T.W.;
RT "Identification of two isoforms of the Cak receptor kinase that are
RT coexpressed in breast tumor cell lines.";
RL Oncogene 12:1469-1477(1996).
RN [6]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Brain and Keratinocytes;
RX MEDLINE=94043265; PubMed=8226977;
RA di Marco E., Cuttuli N., Guerra L., Canceda R., de Luca M.;
RT "Molecular cloning of trke, a novel trk-related putative tyrosine
RT kinase receptor isolated from normal human keratinocytes and widely
RT expressed by normal human tissues.";
RL J. Biol. Chem. 268:24290-24295(1993).
RN [7]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Ovary;
RX MEDLINE=95151638; PubMed=7848919;
RA Laval S., Butler R., Shelling A.N., Hanby A.M., Poulson R.,
RA Ganesan T.S.;
RT "Isolation and characterization of an epithelial-specific receptor
RT tyrosine kinase from an ovarian cancer cell line.";
RL Cell Growth Differ. 5:1173-1183(1994).
RN [8]
RP SEQUENCE FROM N.A.
RA Shima S., Tamiya G., Oka A., Inoko H.;
RT "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Muscle;

```

RA Strausberg R.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND  
 CC RECOGNITION.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, 1/CAK I (SHOWN HERE) AND 2/CAK  
 CC II; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER BY THE  
 CC ABSENCE OF A 37 RESIDUES SEGMENT.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN MOST ADULT TISSUES  
 CC AND IS HIGHEST IN THE BRAIN AND LUNG. ABUNDANT IN BREAST CARCINOMA  
 CC CELL LINES.  
 CC -1- DOMAIN: THE GLY/PRO-RICH DOMAINS MAY BE REQUIRED FOR AN UNUSUAL  
 CC GEOMETRY OF INTERACTION WITH LIGAND OR SUBSTRATES.  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN  
 CC RECEPTOR SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 FS/8 TYPE C DOMAIN.  
 CC -1- DATABASE: NAME=PROW; NOTE=PROW 1:41-46(2000);  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/1986613159\_g.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: L11315; AAA02866.1; -;  
 DR EMBL: L20817; AAA18019.1; -;  
 DR EMBL: U8705; AAC50917.1; -;  
 DR EMBL: X8208; CAA66871.1; JOINED.  
 DR EMBL: X99023; CAA66871.1; JOINED.  
 DR EMBL: X99024; CAA66871.1; JOINED.  
 DR EMBL: X99025; CAA66871.1; JOINED.  
 DR EMBL: X99026; CAA66871.1; JOINED.  
 DR EMBL: X99027; CAA66871.1; JOINED.  
 DR EMBL: X99028; CAA66871.1; JOINED.  
 DR EMBL: X99029; CAA66871.1; JOINED.  
 DR EMBL: X99030; CAA66871.1; JOINED.  
 DR EMBL: X99031; CAA66871.1; JOINED.  
 DR EMBL: X99032; CAA66871.1; JOINED.  
 DR EMBL: X99033; CAA66871.1; JOINED.  
 DR EMBL: X99034; CAA66871.1; JOINED.  
 DR EMBL: L57508; AAB05208.1; -;  
 DR EMBL: X74979; CAA52915.1; -;  
 DR EMBL: Z28093; CAA82335.1; -;  
 DR EMBL: AP000511; BAB63318.1; -;  
 DR EMBL: BC008716; AAH08716.1; -;  
 DR EMBL: BC013400; AAH13400.1; -;  
 DR HSSP: P11362; IFGK.  
 DR Genew: HGNC:2730; DDR1.  
 DR MIM: 600408; -;  
 DR InterPro: IPR000719; Euk.pkinase.  
 DR InterPro: IPR000421; FAS5.C.  
 DR InterPro: IPR002011; RTKkinaseII.  
 DR InterPro: IPR001245; Tyr.pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00754; FS\_F8\_type\_C; 1.  
 DR ProDom: PD000001; Euk.pkinase; 1.  
 DR SMART: SM00231; FAS5C; 1.  
 DR SMART: SM00219; TYK6; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE\_NEG.  
 DR PROSITE: PS00109; PROTEIN KINASE\_TYR; 1.  
 DR PROSITE: PS00233; RECEPTOR\_TYR\_KIN\_II; 1.  
 DR PROSITE: PS01285; FAS5C\_1; 1.  
 DR PROSITE: PS01286; FAS5C\_2; 1.  
 DR Transferase; Tyrosine-protein kinase; Glycoprotein; Signal;  
 KM Phosphorylation; Transmembrane; Receptor; ATP-binding; Antigen;  
 KW Alternative splicing.  
 FT SIGNAL 1 18 POTENTIAL.

FT	CHAIN	19	913	EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1.
FT	DOMAIN	416	416	EXTRACELLULAR (POTENTIAL).
FT	TRANSLEM	417	443	POTENTIAL.
FT	DOMAIN	444	913	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	31	185	FS/8 TYPE C (PHOSPHOLIPID-BINDING, POTENTIAL).
FT	DOMAIN	377	415	GLY/PRO-RICH.
FT	DOMAIN	476	601	GLY/PRO-RICH.
FT	DOMAIN	610	905	PROTEIN KINASE.
FT	MOD BIND	616	624	ATP (BY SIMILARITY).
FT	BINDING	655	655	ATP (BY SIMILARITY).
FT	ACT SITE	766	766	BY SIMILARITY.
FT	DISULFID	31	185	BY SIMILARITY.
FT	MOD RES	513	513	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	792	792	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	796	796	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	211	211	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	371	371	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	394	394	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	506	542	MISSING (IN ISOFORM 2).
FT	CONFLICT	94	94	L -> V (IN REF. 2 AND 3).
FT	CONFLICT	833	833	L -> V (IN REF. 2 AND 3).
FT	CONFLICT	847	867	OUTDEVIENAGEFRDGRQ -> SAHRBGRHRENGVLP GPGPA (IN REF. 4).
FT	SEQUENCE	913 AA;	101127 MW;	C96913EA906C481E CRC64;

Query Match 7.9%; Score 84; DB 1; Length 913;  
 Best Local Similarity 22.7%; Pred. No. 4.1; Mismatches 73; Indels 40; Gaps 7;  
 Matches 39; Conservative 20;

Qy 7 NOTSPGATEDSYSGWYIDEPQ---GGXELQEGEVPSCHT-SIPPGLYHACLASISIL 61  
 Db 370 NNSPALGSGTPPAPMWPGPPTNFSLELEPRGQPAKAGSPTALIGCLVAIILL 429  
 Qy 62 VILLAMLV-----RRQLWPD-----CVRGPGGLSPVDL----- 93  
 Db 430 LLITLMLMLRLHMRRLSKAKRVLEELTVHLSVPGDTILINRFGREPPYQGEPRP 489  
 Qy 94 AGDRPAPVAPVAVFWLSSICLLLPDBDALPFTLTASAPSDGKTEAPPGAW 145  
 Db 490 RGNPHSAPC---VPGNSALLSNPARYRLATYARPPGPG---PPPAW 534

RESULT 4  
 CEBSR\_RAT STANDARD; PRT; 281 AA.  
 AC P56261;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE CCNAT/enhancer binding protein epsilon (C/EBP epsilon) (C/EBP-related  
 DE protein 1)  
 GN CEBSR OR CREP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98256268; PubMed=9593684;  
 RA Williams S.C., Du Y., Schwartz R.C., Weller S.R., Orlie M.,  
 RA Kettler U.R., Johnson P.F.;  
 RT "C/EBPepsilon is a myeloid-specific activator of cytokine, chemokine,  
 RT and macrophage-colony-stimulating factor receptor genes";  
 RL J. Biol. Chem. 273:13493-13501(1998).  
 RN [2]  
 RP SEQUENCE OF 31-281 FROM N.A.  
 RX MEDLINE=91357471; PubMed=1884998;  
 RA Williams S.C., Cantwell C.A., Johnson P.F.;  
 RT "A Family of C/EBP-related proteins capable of forming covalently  
 RT linked leucine zipper dimers in vitro";

RL Genes Dev. 5:1553-1567(1991).  
 CC -1- FUNCTION: C/EBP ARE DNA-BINDING PROTEINS THAT RECOGNIZE TWO  
 CC DIFFERENT MOTIFS: THE CCAAT HOMOLOGY COMMON TO MANY PROMOTERS AND  
 CC THE ENHANCED CORE HOMOLOGY COMMON TO MANY ENHANCERS (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: BINDS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS  
 CC WITH C/EBP DELTA.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP  
 CC PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF034716; AAC24455.1; -  
 CC InterPro: IPR004827; TF\_ZBP1.  
 CC Pfam: PF00170; BZIP\_1.  
 CC SMART: SM00338; BRLZ; 1.  
 CC Transcription regulation; Activator; DNA-binding; Nuclear protein.  
 CC FT DNA\_BIND 208 228 BASIC MOTIF.  
 CC FT DOMAIN 239 267 LEUCINE-ZIPPER.  
 CC FT SEQUENCE 281 AA; 30589 MW; 68816455C034E250 CRC64;  
 CC -----  
 CC Query Match 7.8%; Score 83; DB 1; Length 281;  
 CC Best Local Similarity 28.1%; Pred. No. 1.5; Indels 46; Gaps 11;  
 CC Matches 50; Conservative 17; Mismatches 65;  
 CC -----  
 CC QY 18 SYGMYIDEPOGKXELDP-----EGEV-PSGHTSIPPLGYHACLASISLV----- 62  
 CC DB 2 SHGYTYCEPFGCG--QPLEFSGRAGPELDWCE-----HEASIDSAIVEGEE 51  
 CC QY 63 LLLIAMVRRRLQMDPCVGRGRLSPVDPLAGD-RPRAVPAVFWLLSLCLLPD 121  
 CC DB 52 QLSLDLQAMKPTPEASLKK-PETSPFHYLPADRPAPVPSHTFG-----PDRX 100  
 CC QY 122 ALPPLTLASAPSODGKT---EAPRG-AWKILGLFHYALYPLAACATAGTAHL 173  
 CC DB 101 ALGPIVSNQSYDPAVAVVKEPRGEGNRTGSGYNPLQYQVANC---GQAVVHL 155  
 CC -----  
 CC RESULT 5  
 CC NCAP\_PHV STANDARD: PRT; 433 AA.  
 CC AC P20247;  
 CC DT 01-AUG-1991 (Rel. 19, Created)  
 CC DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Nucleocapsid protein (Nucleoprotein).  
 CC GN S.  
 CC OS Prospect Hill virus (PHV).  
 CC OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Hantavirus.  
 CC NX NCBI\_TaxID=11603;  
 CC RX [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=PHV-1;  
 CC RA MEDLINE=90177212; PubMed=2309440;  
 CC RA "Partridge M.A., Kang C.Y.;  
 CC RT "Nucleotide sequence analysis of the S genomic segment of Prospect  
 CC RT Hill virus: comparison with the prototype Hantavirus";  
 CC RT Virology 175:167-175(1990).  
 CC CC -1- SUBCELLULAR LOCATION: INTERNAL PROTEIN OF VIRUS PARTICLE.  
 CC -1- SIMILARITY: TO OTHER HANTAVIRUS S SEGMENT PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: M34011; AAA47086.1; -  
 CC DR PIR: A34681; VHVUHP.  
 CC DR InterPro: IPR002214; Hanta\_nucleocap.  
 CC DR Pfam: PF00846; Hanta\_nucleocap; 1.  
 CC DR ProDom: PD001501; Hanta\_nucleocap; 1.  
 CC FT Nucleocapsid; RNA-binding.  
 CC FT CONFLICT 7 7 I -> T (IN REF. 1; AAA47086).  
 CC FT SEQUENCE 433 AA; 49005 MW; EBBDB161AD32A0CD CRC64;  
 CC -----  
 CC Query Match 7.5%; Score 80; DB 1; Length 433;  
 CC Best Local Similarity 24.1%; Pred. No. 4.4;  
 CC Matches 35; Conservative 21; Mismatches 55; Indels 34; Gaps 6;  
 CC -----  
 CC QY 23 YIDPQGGKELQPGVPSCHTSIPPLGYHACLASISLVLL-----LMLVRRRLQMDP 78  
 CC DB 178 YVSMPTNOSTMKABEELTPGFRTIVCGLPQAIWARIISPMVGIGFAFVKD---WAD 234  
 CC QY 79 CVRG-----RPGLPSP-VVDPLAGDRPRAVPAVFWLLSLCLL-----LPDE 120  
 CC DB 235 KVAAPFDQKCPFKAEPRPGPAGAEPLSSIR-----AYLMNRQAVDELTHLPDI 285  
 CC QY 121 DALPPLTLASAPSODGKTAPRGAN 145  
 CC DB 286 DALVELAASGDPTLPDLSLENPLAM 310  
 CC -----  
 CC RESULT 6  
 CC NFPC2\_HUMAN STANDARD: PRT; 925 AA.  
 CC AC Q13469; Q13468; Q9UR2; Q9NPX6; Q9N0H3;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Nuclear factor of activated T-cells, cytoplasmic 2 (T cell  
 CC transcription factor NFAT1) (NFAT pre-existing subunit) (NF-ATp).  
 CC GN NFATC2 OR NFAT1 OR NFATP.  
 CC OS Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NX NCBI\_TaxID=9606;  
 CC RX [1]  
 CC RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 CC RC MEDLINE=96251346; PubMed=8668213;  
 CC RA Luo C., Burgeon E., Carew J.A., McCaffrey P.G., Badalian T.M.,  
 CC RA Lane W.S., Hogan P.G., Rao A.;  
 CC RT "Recombinant NFAT1 (NFATp) is regulated by calcineurin in T cells and  
 CC RT mediates transcription of several cytokine genes";  
 CC RT Mol. Cell. Biol. 16:3955-3966(1996).  
 CC RN [2]  
 CC -----  
 CC RP SEQUENCE FROM N.A.  
 CC RC MEDLINE=21638749; PubMed=11780052;  
 CC RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 CC RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,  
 CC RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 CC RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 CC RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 CC RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 CC RA Clegg S., Coley V.E., Collier R.E., Connor R.B., Corby N.R.,  
 CC RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 CC RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 CC RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 CC RA Hammond S., Harley J.L., Heath P.D., Ho S., Howden J.L., Howden P.J.,  
 CC RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 CC RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 CC RA Leharajaho M.H., Leverhna M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 CC RA Marsh V.L., Martin S.L., McCormack L.J., McEay K., McNurray A.A.,  
 CC RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 CC RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 CC RA Phillimore B.J.C.T., Prachinang S.R., Plumb R.W., Ramsey H.,  
 CC RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tremains A.C., Vaudin M., Walli M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbini R.M., Bentley D.R., Beck S.,  
 RA Rogers J.,  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN  
 RP REVIEW.  
 RX MEDLINE:99189746; PubMed:10089876;  
 RA Crabtree G.R.; "Genetic signals and specific outcomes: signaling through Ca2+;  
 RT calcineurin, and NF-AT.";  
 RL Cell 96:611-614(1999).  
 CC -1- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE  
 CC GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2, IL-3,  
 CC IL-4, TNF-ALPHA OR GM-CSF.  
 CC -1- SUBUNIT: MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX  
 CC THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING  
 CC CYTOPLASMIC COMPONENT NFATC2 AND AN INDUCIBLE NUCLEAR COMPONENT  
 CC NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE  
 CC ACTIVATING PROTEIN-1 FAMILY, MAF, SATM4 AND C/EBP300 CAN ALSO BIND  
 CC THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND  
 CC NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-  
 CC MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT  
 CC TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED  
 CC AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF  
 CC NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: B AND C (SHOWN HERE);  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THYMUS, SPLEEN, HEART, TESTIS,  
 CC BRAIN, PLACENTA, MUSCLE AND PANCREAS.  
 CC -1- INDUCTION: INDUCIBLY EXPRESSED IN T LYMPHOCYTES UPON ACTIVATION OF  
 CC THE T-CELL RECEPTOR (TCR) COMPLEX. INDUCED AFTER CO-ADDITION OF  
 CC PHORBOL 12-MYRISTATE 13-ACETATE (PMA) AND IONOMYCIN.  
 CC -1- DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND  
 CC COOPERATIVE INTERACTIONS WITH API FACTORS (BY SIMILARITY).  
 CC -1- PTM: PHOSPHORYLATED BY NFATC-KINASE, DEPHOSPHORYLATED BY  
 CC CALCINEURIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC -----  
 CC EMBL: U43341; AAC50886.1; -;  
 CC EMBL: U43342; AAC50887.1; -;  
 CC EMBL: AL035682; CAC00529.1; -;  
 CC EMBL: AL035682; CAC054871.1; -;  
 CC EMBL: AL132866; CAC00528.1; -;  
 CC TRANSFAC: T01948; -;  
 CC Genew; HGNC:7776; NFATC2.  
 CC MIM: 600490; -;  
 CC InterPro; IPR002909; IPT\_TIG;  
 CC InterPro; IPR000451; NF\_Rel\_dor\_fam.  
 CC Pfam; PF00554; RHD; 1.  
 CC Pfam; PF01833; TIG; 1.  
 CC SMART; SM00429; TIG; 1.  
 CC PROSITE; PS01204; REL\_1; FALSE\_NEG.  
 CC PROSITE; PS00554; REL\_2; 1.  
 CC Transcription regulation; Activator; Nuclear protein; DNA-binding;  
 CC Transcription regulation; Repeat.  
 CC Alternative splicing; Phosphorylation; Repeat.  
 CC DOMAIN 111 116 CALCINEURIN-BINDING  
 CC FT DOMAIN 119 199 TRANS-ACTIVATION DOMAIN A (TAD-A).  
 CC FT DOMAIN 184 286 3 X APPROXIMATE SP REPEATS.  
 CC FT REPEAT 184 200 1.  
 CC FT REPEAT 213 229 2.

FT REPEAT 272 286 3 (APPROXIMATE).  
 FT DOMAIN 251 253 NUCLEAR LOCALIZATION SIGNAL.  
 FT DOMAIN 421 428 DNA-BINDING.  
 FT DOMAIN 664 666 NUCLEAR LOCALIZATION SIGNAL.  
 FT DOMAIN 904 913 NUCLEAR EXPORT SIGNAL.  
 FT MOD RES 110 110 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARSPIC 908 925 VNEIRKFEFGPPARQOT -> ELIDTILSWIONIL (IN  
 FT ISOFORM B).  
 FT CONFLICT 65 65 M -> L (IN REF. 2).  
 FT SEQUENCE 925 AA; 100163 MW; A825164A9C279025 CRC64;  
 SQ  
 Query Match 7.5%; Score 80; DB:1; Length 925;  
 Best local similarity 23.5%; Pred. No. 9.7;  
 Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;  
 Oy 3 OPAGNOTSPGATEDYSGSMYIDEPQGX-----ELQPGGV-PSCHTSIPPLGYHAC 54  
 Db 240 RPARRSSPPKARRHSGAEALVALPGASPGQRSPSPQPSHVAPDDHGS-PAG--YRP 296  
 Oy 55 IASILVLLMLAVR-----RLQV-----PDGVRGRP---GLP---SPYDPLA--- 94  
 Db 297 VAGSAVITMDALNSLSDSPGCIIPKMKTSPPSPVSAAPSKALPRHIVPAVEFLGPCE 356  
 Oy 95 -GDRPAVPAVFWVLSSLCLLPDEDLPLTL---ASAP-----SQDG----- 136  
 Db 357 QGERNSAPESILVLPYTPKPLVP--AIPICSIPTVASLPLEWPLSSQSGSYELRIE 413  
 Oy 137 -----KTEAPGAWKIGLFHYAALVYPLAACAGTAAHLGSLTSMWHLGV 185  
 Db 414 VQPKPHARAYETEGSGAVK-----APTGHVPVQVHG- YMKRPLGL 456  
 Oy 186 QVW 188  
 Db 457 QIF 459  
 RESULT 7  
 ID DDRI\_RAT STANDARD; PRT; 910 AA.  
 AC 063474;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Bipitelial discoidin domain receptor 1 precursor (EC 2.7.1.112)  
 DE (Tyrosine-protein kinase CAK) (Cell adhesion kinase) (Tyrosine kinase  
 DE DDR) (Discoidin receptor tyrosine kinase) (Protein-tyrosine kinase  
 DE PTK-3).  
 GN DDRI OR EDDRI OR PTK3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPRAIN-Sprague-Dawley; TISSUE=Brain;  
 RC MEDLINE:94173920; PubMed:8127887.  
 RA Sanchez M.P., Tagley P., Saini S.S., He B., Pulido D., Barbacid M.;  
 RT "Multiple tyrosine protein kinases in rat hippocampal neurons:  
 RT Isolation of Ptk-3, a receptor expressed in proliferative zones of  
 RT the developing brain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1819-1823(1994).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND  
 CC RECOGNITION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: VARIOUS EMBRYONIC AND ADULT TISSUES; ALSO  
 CC PROLIFERATIVE ZONES OF THE DEVELOPING BRAIN; HIPPOCAMPAL NEURONS.  
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF TYR-PROTEIN  
 CC KINASES.  
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN  
 CC RECEPTOR SUBFAMILY.



```

DR Pfam: PF00017, SH2, 1.
DR Pfam: PF00169, PH, 1.
DR Pfam: PF00788, RA, 1.
DR PRINTS: PR00401, SH2DOMAIN.
DR Prodom: PD000093, SH2, 1.
DR SMART: SM00233, PH, 1.
DR SMART: SM00314, RA, 1.
DR SMART: SM00252, SH2, 1.
DR PROSITE: PS50003, PH_DOMAIN, 1.
DR PROSITE: PS50001, SH2, 1.
DR SH2 domain; Phosphorylation; Alternative splicing.
FT DOMAIN 229 338
FT DOMAIN 431 512
FT CONFLICT 18 18 W -> C (IN REF. 4).
SQ SEQUENCE 532 AA; 59764 MW; A68679F83A146F74 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 532;
Best Local Similarity 28.3%; Pred. No. 7.4;
Matches 39; Conservative 12; Mismatches 62; Indels 25; Gaps 7;

Qy 75 LMPDCVRGRGRLPSVDF-LAGDRPRAVPAVFMVLSLCLLPDDEL- P 124
Db 17 LMP-ACGTGPRPRPDPLPEVGRSP-----LLPTGRLKEERRATSLPSIPMP 70
Qy 125 FLTLASPSQDGKTEAPRCAMKIL---GLHYVALTYPLAC---ATGHTAAHLGS 176
Db 71 FPELCGPPSPGSLGPGSPSARGLLPRDASRPHVVKVYSEDCAGSVEVAGATARHVCEN 130
Qy 177 TLSSMAH-LGVQVWQRAEC 193
Db 131 LVORAHALSDETWGLVEC 148

RESULT 9
CCBS OENBE STANDARD; PRT; 577 AA.
ID CCBS OENBE STANDARD; PRT; 577 AA.
AC 004648;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Probable cytochrome c biosynthesis protein.
OS Oenothera biennis (Bertero's evening primrose).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Myrtales; Onagraceae; Oenothera.
OC NCBI_TaxID=3950;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93288005; PubMed=8389979;
RX Schuster W., Combettes B., Flieger K., Brennicke A.;
RT "A plant mitochondrial gene encodes a protein involved in cytochrome
RT c biogenesis.";
RL Mol. Genet. 239:49-57(1993).
CC -1- FUNCTION: COULD BE INVOLVED IN ASSEMBLY AND MATURATION OF
CC CYTOCHROMES C. MAY PLAY A ROLE IN GUIDANCE OF APCYCCHROMES AND
CC HEME GROUPS FOR THE COVALENT LINKAGE INTRODUCED BY THE CYTOCHROME-
CC C-HEME LIASE.
CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCLL/NREB/CCSA FAMILY.
CC -1- CAUTION: EXTENSIVELY MODIFIED BY RNA-EDITING. WE HAVE ONLY
CC MODIFIED THE GENOMIC SEQUENCE FOR THE OBSERVED R->W EDITING.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X69555; CAA49287.1; ALT SEQ.
CC InterPro: IPR002541; CytC_aem.
CC InterPro: IPR003567; Cyt_c_bioG.

```

```

DR Pfam: PF01578; CytC_aem; 2.
DR PRINTS: PR01410; CCBIOGENESIS.
DR Mitochondrion; Cytochrome c-type biogenesis; RNA editing.
DR SEQUENCE 577 AA; 65238 MW; 10PB05972BBA8B CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 577;
Best Local Similarity 24.5%; Pred. No. 8.1;
Matches 27; Conservative 18; Mismatches 46; Indels 19; Gaps 4;

Qy 13 ATEYSGSWYIDPQGXLEIPEG-EVPSCHTSLPPGYACASLIVLLAMLR 71
Db 438 AHHELGGWGMFMDPVENASFMRVUATARIHSVLP-LLNSCHSLNIVTLIC----- 490
Qy 72 RROLWPDGVRG---RPGLPSVDFLAGDRPRAVPAVFMVLSLCLL 117
Db 491 -----CVLGTFSIRGLASVHSFATDTRGIFLMRPFLLMTGISMIT 533

RESULT 10
OL56_STRAT STANDARD; PRT; 3519 AA.
ID OL56_STRAT STANDARD; PRT; 3519 AA.
AC 007017;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oleandomycin polyketide synthase, modules 5 and 6.
DE ORP.
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteridae (class); Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OC NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94150470; PubMed=8107683;
RX Swan D.G., Rodriguez A.M., Vilches C., Salas J.A.;
RT "Characterisation of a Streptomyces antibioticus gene encoding a type
RT I polyketide synthase which has an unusual coding sequence.";
RL Mol. Genet. 242:358-362(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC LACTONE RING.
CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: L09654; AAA19695.1; -.
CC HSSP: P25715; TMA.
DR InterPro: IPR001227; Acyltransferase.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; Pantine_attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00109; ketoacyl-synt. 2.
DR Pfam: PF00550; po-binding; 2.
DR Pfam: PF00698; Acyl trans; 2.
DR Pfam: PF00975; Thioesterase; 1.
DR Pfam: PF02801; ketoacyl-synt C; 2.
DR PROSITE: PS00612; PHOSPHOPANTETHEINE; 2.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 2.
DR Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine; Multifunctional enzyme; Repeat.
FT DOMAIN 1 3519
FT DOMAIN 32 501
FT DOMAIN 569 880
FT DOMAIN 1200 1382
FT DOMAIN 1487 1561
FT DOMAIN 1487 1561

```



```

FT DOMAIN 1686 2156 BETA-KETOACYL SYNTHASE 2.
FT DOMAIN 2220 2541 ACYLTRANSFERASE (AT) 2.
FT DOMAIN 2856 3038 BETA-KETOACYL REDUCTASE 2.
FT DOMAIN 3141 3215 ACYL CARRIER (ACP) 2.
FT ACT SITE ? ? ACYL-ENZYME INTERMEDIATE.
FT BINDING ? ? PHOSPHOPANTETHEINE (BY SIMILARITY).
FT DOMAIN 3270 3519 ? ? THIOESTERASE.
FT ACT SITE 210 210 BETA-KETOACYL SYNTHASE.
FT ACT SITE 660 660 ACYL-ENZYME INTERMEDIATE.
FT BINDING 1203 1249 NADP.
FT ACT SITE 1524 1524 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACT SITE 1859 1859 BETA-KETOACYL SYNTHASE.
FT ACT SITE 2311 2311 ACYL-ENZYME INTERMEDIATE.
FT NP BIND 2859 2905 NADP.
FT BINDING 3178 3178 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 3519 AA; 368561 MW; 41AE7BAABAE61P86 CRC64;

Query Match 7.4%; Score 78.5; DB 1; Length 3519;
Best Local Similarity 23.5%; Pred. No. 54;
Matches 54; Conservative 25; Mismatches 70; Indels 81; Gaps 12;

QY 11 PGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLSILLVLLMLV 70
DB 2642 PAALDEVLGCGMLFVVRG-----LADGVAVRVAALVTA 2676
QY 71 R-----RQLWPCVGRPGLPSPVDFLAGDRR-----AVRA--AVRMVL 109
DB 2677 RCGEVSVELDPTPRDBRAVAEVAAGR-GVSGVSTLSMDRRHSEHVAVGASLVL 2735
QY 110 LSSLC-----LLPDEDLPPLVTLASAPSDCKTEARGA--W--KIIGLPH-- 152
DB 2736 AQAIVLDGKRGVCEGRMLVLTTRDA-----VVAQPSDAGAVTIPVQAGVWGFGRVLEHPE 2790
QY 153 --VAALVYPLAA--CATAGHTAAHLTGSLTSMHL-----GVQWQWR 190
DB 2791 LMGGLIDLPVEAPEPGSTCDHTYADLATVVASAGDEQVAVRSGWVR 2840

RESULT 11
CEBE: HUMAN STANDARD; PRT; 281 AA.
ID CEBE_HUMAN STANDARD; PRT; 281 AA.
AC Q15744; Q15745; Q99803;
DC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE CCAAT/enhancer binding protein epsilon (C/EBP epsilon).
GN CEBPE.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=96299737; PubMed=8661101;
RA Antonson P., Stellan B., Yamanka R., Xanthopoulos K.G.;
RT "A novel human CCAAT/enhancer binding protein gene, C/EBPepsilon, is
RT expressed in cells of lymphoid and myeloid lineages and is localized
RT on chromosome 14q11.2 close to the T-cell receptor alpha/delta
RT locus."
RL Genomics 35:30-38(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97184462; PubMed=9032264;
RA Chumakov A.M., Grillier I., Chumakova E., Chih D., Slater J.,
RA Koefler H.P.;
RT "Cloning of the novel human myeloid-cell-specific C/EBP-epsilon
RT transcription factor."
CC MO1. Cell. Biol. 17:1375-1386(1997).
CC -1- FUNCTION: C/EBP ARE DNA-BINDING PROTEINS THAT RECOGNIZE TWO
CC DIFFERENT MOTIFS: THE CCAAT HOMOLOGY COMMON TO MANY PROMOTERS AND
CC THE ENHANCED CORE HOMOLOGY COMMON TO MANY ENHANCERS.
CC -1- SUBUNIT: BINS DNA AS A DIMER AND CAN FORM STABLE HETERODIMERS
CC WITH C/EBP DELTA.

```

```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Strongest expression occurs in promyelocyte
CC and late-myeloid-like cell lines.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
CC PROTEINS. Ref.2 sequence differs from that shown due to a
CC frameshift in position 4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48865; AAC50708.1; -.
DR EMBL; U48866; AAC50709.1; -.
DR EMBL; U80982; AAC51130.1; ALT_FRAME.
DR TRANSFAC; T04883; -.
DR Genew; HGNC:1836; CEBPE.
DR MIM; 600749; -.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP; 2.
DR SMART; SM00338; BRLZ; 1.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Phosphorylation.
FT DNA BIND 208 228 BASIC MOTIF.
FT DOMAIN 239 267 LEUCINE-ZIPPER.
FT CONFLICT 64 64 P -> S (IN REF. 2).
FT CONFLICT 68 68 R -> G (IN REF. 2).
FT CONFLICT 252 252 Q -> E (IN REF. 2).
SQ SEQUENCE 281 AA; 30702 MW; 2F808717AD5ACD2B CRC64;

Query Match 7.4%; Score 78; DB 1; Length 281;
Best Local Similarity 28.9%; Pred. No. 4.2;
Matches 52; Conservative 16; Mismatches 62; Indels 50; Gaps 11;

QY 18 SYGSWYIDEPQGXELP-----EGEVPSCHTSIPGLYHACLAS-----LSI 60
DB 2 SHCTTYECPRGQ--QLERSGGRACPGELGDCHASIDLSA-----TESGEQLSD 56
QY 61 LVLLILMLVRRRQLPDCVGRGRLPSPVDFLAGD-RPRAVPAVFWLLSLCLLPD 119
DB 57 LRAVKAPEPAR-----LKG-FGTPAPFHYLPDPRPFAVPHPTFG-----PD 98
QY 120 EDALPFLTLASAPSDGKT-----EAPRG--AKIIDLPHYALVYPLAACATAGHTAAH 173
DB 99 RKAIGGTYSSGSDPRAAVAVBERPGRGSAARGSYNVOIVAHG---GOTAAHL 155

RESULT 12
PFTB_PEA STANDARD; PRT; 419 AA.
ID PFTB_PEA STANDARD; PRT; 419 AA.
AC Q04903;
DC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein farnesyltransferase beta subunit (EC 2.5.1.-) (CAAX
DE farnesyltransferase beta subunit) (RAS proteins prenyltransferase
DE beta) (Frase-beta).
GN PFTB.
OS Pisum sativum (garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosoid I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
ON NCBI_TaxID=3888;
RX MEDLINE=94105305; PubMed=8278509;
RA Yang Z., Cramer C.L., Watson J.C.;

```



RT "protein farnesyltransferase in plants. Molecular cloning and  
 RT expression of a homolog of the beta subunit from the garden pea."  
 RL Plant Physiol. 101:667-674 (1993).  
 CC -1- FUNCTION: CATALYZES THE TRANSFER OF A FARNESYL MOIETY FROM  
 CC FARNESYL PYROPHOSPHATE TO A CYSTEINE AT THE FOURTH POSITION FROM  
 CC THE C-TERMINUS OF SEVERAL PROTEINS. THE BETA SUBUNIT IS  
 CC RESPONSIBLE FOR PEPTIDE-BINDING (BY SIMILARITY).  
 CC -1- CORACATOR: BINDS ONE ZINC ION (BY SIMILARITY).  
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PROTEIN PRENYLTRANSFERASE BETA SUBUNIT  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 PPTB REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L08664; AAA33649.1; -  
 CC PIR: J02254; J02254.  
 CC HSSP: Q02293; 1FT1.  
 CC InterPro: IPR001330; Prenyltrans.  
 CC Pfam: PF00432; Prenyltransferase; 5.  
 CC KMW transferase; Prenyltransferase; Repeat; zinc.  
 CC FT REPEAT 68 109 PFTB 1.  
 CC FT REPEAT 119 160 PFTB 2.  
 CC FT REPEAT 167 208 PFTB 3.  
 CC FT REPEAT 215 256 PFTB 4.  
 CC FT REPEAT 329 371 PFTB 5.  
 CC FT REPEAT 241 241 ZINC (BY SIMILARITY).  
 CC FT METAL 243 243 ZINC (BY SIMILARITY).  
 CC FT METAL 359 359 ZINC (BY SIMILARITY).  
 CC SQ SEQUENCE 419 AA; 46793 MW; 4F040E0094277D7C CRC64;  
 CC -----  
 CC Query Match 7.4%; Score 78; DB 1; Length 419;  
 CC Best Local Similarity 19.3%; Pred. No. 6.4;  
 CC Matches 52; Conservative 39; Mismatches 70; Indels 108; Gaps 12;  
 CC -----  
 CC QY 26 EPOGKELPEGE--VPSCHTSPGGLYHACIASLIL--VLLILMLVRRQLMPDCV 80  
 CC DB 131 QPNGGGRMHDEGIDIVACVYAI-----SVASVILNIDDELLKNVDFTLSCQTEGGL 184  
 CC QY 81 RRRPGLPSPVDF-----LAG-----DRPRAVPAVVF----- 106  
 CC DB 185 AGEPGSEAHGCTFGCLAMILIGVKNLDRPRLLDVYFRQGEKCGFGRTNKLVDGCY 244  
 CC QY 107 -----WLLSSLCILLPDE--DALPELTLASAPSQ-----DGKTPA 140  
 CC DB 245 SFMOGGAVALLQRLHSHIIDEQVAEASQFTVSDAEKEXCLDGTSSHATSHIRHEGNTS 304  
 CC QY 141 PRGAWKILG-----LFHYAALVYPLAACATAG-----HTAAHL 173  
 CC DB 305 CSSDVKNIGNTFSEWROSEPLFHSIALQOYLILCSGSDGDLKDKGRKHVHNSCYCL 364  
 CC QY 174 LGSTL---SHALGVQVWQRAECPOVPKI 199  
 CC DB 365 SGLSLCOYSWS-----KRPSFLPKV 386  
 CC -----  
 CC RESULT 13  
 CC DDRI\_MOUSE STANDARD; PRT; 911 AA.  
 CC AC Q03146;  
 CC DT 01-OCT-1994 (Rel. 30, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Epithelial discoidin domain receptor 1 precursor (EC 2.7.1.112)  
 CC DE (Tyrosine-protein kinase CAK) (Cell adhesion kinase) (Tyrosine kinase  
 CC DB) (Discoidin receptor tyrosine kinase) (Protein-tyrosine kinase

DE MPK-6).  
 CN DDRI OR EDDRI OR CAK OR MPK6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 ON NCBI\_TaxId=10090;  
 OX [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=96204002; PubMed=8622863;  
 RA Perez J.L., Jing S.Q., Wong T.W.,  
 RT "Identification of two isoforms of the Cak receptor kinase that are  
 RT coexpressed in breast tumor cell lines."  
 RL Oncogene 12:1463-1477(1996).  
 RN [2]  
 RP SEQUENCE OF 766-822 FROM N.A.  
 RC STRAIN=C57BL/6; Tissue=Embryonic brain;  
 RX MEDLINE=93096484; PubMed=1281307;  
 RA Gilardi-Hebenstreit P., Nieto M.A., Frain M., Mattei M.-G.,  
 RA Chetlier A., Wilkinson D.G., Charney P.,  
 RT "An Eph-related receptor protein tyrosine kinase gene segmentally  
 RT expressed in the developing mouse hindbrain."  
 RL Oncogene 7:2499-2506(1992).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELL-CELL INTERACTIONS AND  
 CC RECOGNITION (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, CAK I (SHOWN HERE) AND CAK II;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ONLY DIFFER BY THE  
 CC ABSENCE OF A 37 RESIDUES SEGMENT.  
 CC -1- TISSUE SPECIFICITY: THE PREDOMINANT ISOFORM CAK I, IS EXPRESSED IN  
 CC DEVELOPING EMBRYO AND ADULT BRAIN; CAK II IS EXPRESSED IN VARIOUS  
 CC EPITHELIAL CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN  
 CC RECEPTOR SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 FS/8 TYPE C DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L57509; AAB05209.1; -  
 CC EMBL: X57240; CAA40516.1; -  
 CC PIR: S30502; S30502.  
 CC HSSP: P00523; 2PTK.  
 CC MGD: MGI:99216; Ddrl.  
 CC InterPro: IPR000719; Euk\_pkinase.  
 CC InterPro: IPR00421; FAS8\_C.  
 CC InterPro: IPR002011; RTKinaseII.  
 CC InterPro: IPR01245; Tyr\_pkinase.  
 CC Pfam: PFO0069; pkinase; 1.  
 CC Pfam: PFO0754; F5\_P8 type C; 1.  
 CC ProDom: PDD00001; Euk\_pkinase; 1.  
 CC SMART: SM00231; FAS8C; 1.  
 CC SMART: SM00219; TyKc; 1.  
 CC PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 CC PROSITE: PS00109; PROTEIN\_KINASE\_TIR; 1.  
 CC PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 CC PROSITE: PS00239; RECEPTOR\_TYR\_KIN\_II; 1.  
 CC PROSITE: PS01285; FAS8C\_1; 1.  
 CC PROSITE: PS01286; FAS8C\_2; 1.  
 CC Transferrase; Tyrosine-protein kinase; Glycoprotein; Signal;  
 CC phosphorylation; Transmembrane; Receptor; ATP-binding;  
 CC Alternative splicing.  
 CC KW SIGNAL 1 19  
 CC FT CHAIN 20 911  
 CC FT DOMAIN 20 414  
 CC FT TRANSMEM 415 441  
 CC POTENTIAL.  
 CC EPITHELIAL DISCOIDIN DOMAIN RECEPTOR 1.  
 CC EXTRACELLULAR (POTENTIAL).

```

FT DOMAIN 442 911 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 32 186 F5/8 TYPE C (PHOSPHOLIPID-BINDING,
FT 379 413 POTENTIAL).
FT DOMAIN 474 599 GLY/PRO-RICH.
FT DOMAIN 608 903 PROTEIN KINASE.
FT NP_BIND 614 622 ATP (BY SIMILARITY).
FT BINDING 653 653 ATP (BY SIMILARITY).
FT ACT_SITE 764 764 ATP (BY SIMILARITY).
FT DISULFID 32 186 BY SIMILARITY.
FT MOD_RES 511 511 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 790 790 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 794 794 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 795 795 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 213 213 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPIC 503 539 MISSING (IN ISOFORM CAK II).
SQ SEQUENCE 911 AA; 101160 MW; DBTFE03DD79510 CRC64;

Query Match 7.4%; Score 78; DB 1; Length 911;
Best Local Similarity 22.1%; Pred. No. 14;
Matches 36; Conservative 23; Mismatches 70; Indels 34; Gaps 7;

OY 13 ATEDYSGSYIDEPC---GGELOREGEVPSCHT-SIPGGLYHACIASLIVLLLA 67
DB 374 SDDTPPAPMPPPPPPTNFFSSLEPRGQVPAKESPAIIIGLIVALLILLILA 433
OY 68 MLV-----RRQLMPD-----CVGRGRLSPVDFLAGRPAVPA 103
DB 434 LMLRLHRLRLSKARRVLEELIVHLSVPGDTLLINRRPGPREPPPY-QEPRPCTTP 492
OY 104 AVFVLLSLCLLPED-ALPFLTASAPSGDKTEAPRGAM 145
DB 493 HSAPCVNGSALLSNPAILATYARPPPG---PPTAM 532

RESULT 14
CCBS_DAUCA STANDARD; PRT; 579 AA.
AC 004647;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable cytochrome c biogenesis protein.
OS Daucus carota (Carrot).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasteridae II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxId=4039;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93288005; PubMed=8389979;
RA Schuster W., Combettes B., Flieger K., Bremicke A.;
RT "A plant mitochondrial gene encodes a protein involved in cytochrome
RT c biogenesis."
RL Mol. Genet. 239:49-57(1993).
CC -1- FUNCTION: COULD BE INVOLVED IN ASSEMBLY AND MATURATION OF
CC CYTOCHROMES C. MAY PLAY A ROLE IN GUIDANCE OF APOCTOCHROMES AND
CC HEME GROUPS FOR THE COVALENT LINKAGE INTRODUCED BY THE CYTOCHROME-
CC C-HEME LIGASE.
CC -1- SIMILARITY: BELONGS TO THE CCMF/CYCK/CCL1/NREE/CCSA FAMILY.
CC -1- CAUTION: EXTENSIVELY MODIFIED BY RNA-EDITING.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; X69554; CAA49286.1; -.
DR PIR; S31568; S31568.
DR InterPro; IPR002541; CYC_C_asm.
DR InterPro; IPR003567; CYC_C_blog.
DR Pfam; PF01578; CYC_C_asm; 2.
DR PRINTS; PR01410; CCEIOGENESIS.
KM Mitochondrion; Cytochrome c-type biogenesis; RNA editing.
SQ SEQUENCE 579 AA; 65691 MW; DB2F8C2AFB12F37 CRC64;

Query Match 7.3%; Score 77.5; DB 1; Length 579;
Best Local Similarity 25.5%; Pred. No. 10;
Matches 28; Conservative 16; Mismatches 47; Indels 19; Gaps 4;

OY 13 ATEDYSGSYIDEPCGGELOREPGEVPSCHTSIPGGLYHACIASLIVLLLA 71
DB 440 AHHELGRGCMWRDPENNSFMRPLATARHSVILP-LHSTWSELVITL----- 490
OY 72 RQLMPDCVYG---RPGLPSPVDFLAGRPAVPAVFWLLSLCLL 117
DB 491 -----PCVSGTSSIRSGILAPVHSFATDTRGIFLWRFLMTGISMIL 535

RESULT 15
PASI_MOUSE STANDARD; PRT; 874 AA.
AC P97481; O08787; O55046;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endothelial PAS domain protein 1 (EPAS1), a transcription factor
DE (HNF1) (HIF-related factor) (HNF).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RC MEDLINE=97152468; PubMed=9000051;
RA Tian H., McKnight S.L., Russell D.W.;
RT "Endothelial PAS domain protein 1 (EPAS1), a transcription factor
RT selectively expressed in endothelial cells."
RL Gene Dev. 11:72-82(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Hypothalamus, and Skeletal muscle;
RX MEDLINE=97272213; PubMed=9113979;
RA Ema M., Taya S., Yokotani N., Sogawa K., Matsuda Y.;
RA Fujii-Kuriyama Y.;
RT "A novel bHLH-PAS factor with close sequence similarity to hypoxia-
RT inducible factor 1alpha regulates the VEGF expression and is
RT potentially involved in lung and vascular development."
RL Proc. Natl. Acad. Sci. U.S.A. 94:4273-4278(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain capillary;
RX MEDLINE=97321546; PubMed=9178256;
RA Flameit J., Froehlich T., von Reutern M., Kappel A., Demert A.,
RA Kisse W.;
RT "HIF, a putative basic helix-loop-helix-PAS-domain transcription
RT factor is closely related to hypoxia-inducible factor-1 alpha and
RT developmentally expressed in blood vessels."
RL Mech. Dev. 63:51-60(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE INDUCTION OF OXYGEN
CC REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE
CC ELEMENT (HRE). REGULATES THE VASCULAR ENDOTHELIAL GROWTH FACTOR
CC (VEGF) EXPRESSION AND SEEMS TO BE IMPLICATED IN THE DEVELOPMENT OF
CC BLOOD VESSELS AND THE TUBULAR SYSTEM OF LUNG. MAY ALSO PLAY A ROLE
CC IN THE FORMATION OF THE ENDOTHELIUM GIVING RISE TO THE BLOOD BRAIN
CC BARRIER. POTENT ACTIVATOR OF THE TIE-2 TYROSINE KINASE EXPRESSION.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER

```

BLHLPROTEIN. HETERODIMER WITH THE ARNT PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH HIGHEST LEVELS  
 CC IN LUNG, FOLLOWED BY HEART, KIDNEY, BRAIN AND LIVER. PREDOMINANTLY  
 CC EXPRESSED IN ENDOTHELIAL CELLS. ALSO FOUND IN SMOOTH MUSCLE CELLS  
 CC OF THE UTERUS, NEURONS, AND BROWN ADIPOSE TISSUE. HIGH EXPRESSION  
 CC IN EMBRYONIC CHOROID PLEXUS AND KIDNEY GLOMERULI.  
 CC -1- DEVELOPMENTAL STAGE: IN DAY 11 EMBRYO, EXPRESSION IS ALMOST  
 CC EXCLUSIVELY SEEN IN ENDOTHELIAL CELLS OF THE INTERSEGMENTAL BLOOD  
 CC VESSELS SEPARATING THE SOMITES, THE ATRIAL AND VENTRICULAR  
 CC CHAMBERS OF THE HEART, AND THE DORSAL AORTA. HIGH EXPRESSION ALSO  
 CC OCCURS IN EXTRA-EMBRYONIC MEMBRANES. IN THE DEVELOPING BRAIN OF  
 CC DAY 13 EMBRYO, ENDOTHELIAL CELLS OF THE HIGHLY VASCULARIZED  
 CC CHOROID PLEXUS CONTAIN HIGH LEVELS OF EPAS1.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLL) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-bb.ch](mailto:license@sib-bb.ch)).  
 CC -----  
 DR EMBL; U01983; AAB14496.1; -;  
 DR EMBL; D89787; BAA20130.1; -;  
 DR EMBL; AF045160; AAC12871.1; -;  
 DR TRANSFAC; T02719; -;  
 DR MGD; MGI:109169; Epas1.  
 DR InterPro; IPR001092; HLH basic.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR Pfam; PF00969; PAS; 2.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR TIGRFAMs; TIGR00229; sensory\_box; 2.  
 DR PROSITE; PS00038; HLH; 1.  
 DR PROSITE; PS50112; PAS; 2.  
 KM Repeat, DNA-binding; Nuclear protein; Transcription regulation;  
 KM Activator; Angiogenesis; Developmental protein.  
 FT DNA\_BIND  
 FT 15  
 FT DOMAIN 28  
 FT 154  
 FT DOMAIN 230  
 FT 300  
 FT DOMAIN 304  
 FT 347  
 FT DOMAIN 471  
 FT 479  
 FT CONFLICT 191  
 FT 191  
 FT CONFLICT 439  
 FT 440  
 FT CONFLICT 463  
 FT 463  
 FT CONFLICT 654  
 FT 654  
 FT CONFLICT 663  
 FT 663  
 FT CONFLICT 669  
 FT 669  
 FT CONFLICT 673  
 FT 673  
 FT CONFLICT 678  
 FT 678  
 FT CONFLICT 725  
 FT 725  
 FT CONFLICT 731  
 FT 731  
 FT CONFLICT 762  
 FT 762  
 FT CONFLICT 786  
 FT 786  
 FT CONFLICT 791  
 FT 791  
 FT CONFLICT 794  
 FT 794  
 SQ SEQUENCE 874 AA; 96712 MW; A6FFA90A843640C CRC64;  
 Query Match 7.2%; Score 76.5; DB 1; Length 874;  
 Basic Local Similarity 22.9%; Freq. No. 19;  
 Matches 61; Conservative 17; Mismatches 93; Indels 95; Gaps 14;  
 10 SPG-ATPDVYGSWYIDEPQG-XGLVDPQGE-----VP-----40

Db	406	TTGGATTTSLDFGQNFDEPSAYKALILPQGPWVSGLSRSHSQSESGSLPAFTVPQADTP	465
Qy	41	-----SCHTSIRPGLYHACLAS-LSILYL-LTLAMLYRR-----	73
Db	466	GNTTTPASSSSSCSCTPSSPEDTYSLSLEPLKLEIVIEKTLFAMOTERDGGSTQTFSELDL	535
Qy	74	-----QLMPDCYRGRPGPLSPVDFLACDRPRAVPAALFVWLLSLCLLPD	119
Db	526	ETLAPYIPMDGEDFQLSPDIC-----PEEELMESESQPTPOHCFTSTMSIQPLTPG	576
Qy	120	EDALPELTLASAPQ--GKTEADPRGAMKILGFHYAALYYPLAACATVAGHTAAHLILG--	175
Db	577	ATNGPFF-LDKYPOQLSEKRTSESH--WPMSSIFPDAGSKGSLSPCCGQASTPLSSMGR	633
Qy	176	STLSW-----AHLGVQW-----QRAE	192
Db	634	STNTQMPDPLPIHFGTKMPVGDQSAE	659

Search completed: March 13, 2003, 11:30:20  
Job time : 10.4561 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 13, 2003, 11:27:17 ; Search time 11.3931 Seconds  
(without alignment)  
1679.148 Million cell updates/sec

Title: US-09-816-653A-2

Perfect score: 1061  
Sequence: 1 MSOPAGNQTSPGATEDYSYC.....MAHLGVQVWRAECPOVPK1 199

Scoring table: BLOSUM62

GAPOP 10.0 , GAPEXT 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:  
2: p1r2:  
3: p1r3:  
4: p1r4:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86.5	8.2	757	2	Beta-N-acetylhexos
2	85.5	8.1	380	2	glycoprotein I - b
3	84.5	8.0	335	2	probable anthranil
4	84	7.9	913	2	receptor tyrosine
5	83	7.8	1172	2	probable ABC-type
6	80.5	7.6	395	2	probable MFS trans
7	80	7.5	433	1	nucleosidase prote
8	80	7.5	876	2	transcription fact
9	80	7.5	921	2	type IV prepilin p
10	79.5	7.5	249	1	cytochrome c-type
11	79.5	7.5	577	2	virb6 homolog - Bo
12	79	7.4	463	2	tyrosine kinase re
13	79	7.4	910	2	probable MFS trans
14	78.5	7.4	477	2	epidermal growth f
15	78.5	7.4	532	2	polyketide synthas
16	78.5	7.4	3519	2	larnesyl-diphospha
17	78	7.4	419	2	cytochrome c-type
18	77.5	7.3	519	2	thiamine transport
19	77	7.3	521	2	protein-tyrosine k
20	76.5	7.2	221	2	probable 3-hydroxy
21	76	7.2	394	2	hypothetical prote
22	76	7.2	666	2	hypothetical prote
23	76	7.2	709	2	hypothetical prote
24	75.5	7.1	400	2	cyclin Bb-II - ye
25	75.5	7.1	454	2	L-iduronidase (EC
26	75.5	7.1	655	1	cytochrome c bioge
27	75	7.1	211	2	Lept170_F2_64 pro
28	75	7.1	230	2	cp1 protein - mai
29	75	7.1	668	2	T01685

30	74.5	7.0	425	2	C87619	hypothetical prote
31	74	7.0	320	2	A82449	conserved hypochet
32	73.5	6.9	233	2	T47136	hypothetical prote
33	73.5	6.9	300	2	D83487	hypothetical prote
34	73.5	6.9	636	2	UN0047	class I cytokinase
35	73.5	6.9	1174	2	AH0663	probable pyruvate
36	73.5	6.9	1201	2	A83007	hypothetical prote
37	73	6.9	411	2	D72588	hypothetical prote
38	73	6.9	473	2	T03611	cyclin, B-type - C
39	73	6.9	633	2	P84564	probable protein k
40	73	6.9	739	2	A83015	primosomal protein
41	73	6.9	766	2	F70735	hypothetical prote
42	72.5	6.8	176	2	B83199	hypothetical prote
43	72.5	6.8	258	2	S75950	hypothetical prote
44	72.5	6.8	529	2	C70545	hypothetical prote
45	72.5	6.8	636	1	VCWFS	env polyprotein -

#### ALIGNMENTS

##### RESULT 1

F87304 beta-N-acetylhexosaminidase, probable [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: F87304

R:Hieman, W.C.; Paladlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87304

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-757 <STO>

A:Cross-references: GB:A8005673; NID:g13421618; PIDN:AKK2434.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0447

Query Match	8.2%	Score 86.5	DB 2	Length 757
Best Local Similarity	30.6%	Pred. No. 3.4		
Matches	49	Conservative 16	Mismatches 60	Indels 35
Gaps	9			
Oy	62	VLLLLMLVRR-RQLMPDCAGRGRLSPVDFLAGDPP-----RAVPA 103		
Db	65	VALQLADLLIQRSGFRPRKVEGPPAADIV--LTRBPAGEAYKLDINAKGATIAAKRA 122		
Oy	104	AVFMVLLSSICLLPDEDALPFLTLASAPSDGKTEAPRGAMKILGFHYALYY----P 159		
Db	123	GLTVGAMSLWQLATPDEAKGPVALLAASIE----DAPFRAMR--GLMVDASAHYOSLTD 175		
Oy	160	LAACATAGTAAHLIGSTLSNAHLGVQVWRAECPOVPK1 199		
Db	176	LKAVIDA--MAAHKL-NTFMHMLVDDQGM-RLKIKTKPKL 211		

##### RESULT 2

S35785 glycoprotein I - bovine herpesvirus 1

C:Species: bovine herpesvirus 1

C:Date: 09-Jun-1994 #sequence\_revision 12-May-1995 #text\_change 24-Nov-1999

C:Accession: S35785

R:Audonnet, J.

A:Reference number: S35782

A:Accession: S35785

A:Molecule type: DNA

A:Residues: 1-380 <NUD>

A:Cross-references: EMBL:223068; NID:g312185; PIDN:CAA80605.1; PID:g312189

C:Superfamily: pseudorabies virus glycoprotein gp63

C:Keywords: glycoprotein

Query Match 8.1%; Score 85.5; DB 2; Length 380;

Best Local Similarity 27.6%; Pred. No. 2.1; Matches 37; Conservative 12; Mismatches 39; Indels 47; Gaps 6;

4 PAGNOTSGATEDSYSGWYIDEPQGXELQPEGEVPSCHTSPGLHACIAS-----LS 59

Db 256 PAGPTASPAASPSAFA-----AAPAAAGAPAGDTPA-----RRRLASLIVPLC 302

Qy 60 ILVLLIAM-----LVRRRL-----WPCVGRGRLSPV 90

Db 303 VLVLILLALCATATNCLRRRLPCSRRYKPTTCACGSGTCACRPPCRGAASAPATV 362

Qy 91 DFLAGDRPRAVPA 104

Db 363 VAL-GSRPKAPPLA 375

RESULT 3

T36304 probable anthranilate phosphoribotransferase - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000

C/Accession: T36304

R/Sanders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, March 1999

A/Reference number: Z21604

A/Accession: T36304

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-335 <SAU>

A/Cross-references: EMBL:AL035654; PIDN:CAB3853.1; GSPDB:GN00070; SCODEB:SCDB.05C

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: trpD2; SCODEB:SCDB.05C

C/Superfamily: anthranilate phosphoribosyltransferase; trpD homology

Query Match 8.0%; Score 84.5; DB 2; Length 335;

Best Local Similarity 30.8%; Pred. No. 2.2; Matches 36; Conservative 10; Mismatches 42; Indels 29; Gaps 5;

Qy 81 RGRPGLPSPVDFLAGDRPRAVPAVFMVLSLCLLPDDELPLFLTLASPSODGTEA 140

Db 247 RSRRG-----DLAGGD--RAVMAAVLRVLAG-----GGGRADVVLN 283

Qy 141 PRGAKLGLFHTAALYPLAACATAGTAHLGSLTSMHLCGVQVQRAECPQVP 197

Db 284 AAALRVAGVATWSGDLRLAASAVDGAAGGLD---RMAHAS---WQRADEVVP 334

RESULT 4

A48280 receptor tyrosine kinase - human

C/Species: Homo sapiens (man)

C/Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 24-Sep-1999

C/Accession: A48280

R/Johnson, J.D.; Edman, J.C.; Rutter, W.J.

Proc. Natl. Acad. Sci. U.S.A. 90, 5677-5681, 1993

A/Title: A receptor tyrosine kinase found in breast carcinoma cells has an extracellular

A/Reference number: A48280; MUID:99296201; PMID:8390675

A/Accession: A48280

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-913 <RES>

A/Cross-references: GB:L11315; NID:9403386; PIDN:AA02866.1; PID:9403387

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; discoidin I amino-ter

C/Keywords: ATP

F/30-185/Domain: discoidin I amino-terminal homology <DN1>

F/608-912/Domain: protein kinase homology <KIN>

F/616-624/Region: protein kinase ATP-binding motif

Query Match 7.9%; Score 84; DB 2; Length 913;

Best Local Similarity 22.7%; Pred. No. 7.1; Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;

Qy 7 NOTSPGATEDESYSGWYIDEPD-----GGKELQPEGEVPSCHT-SIPGLHACIASLSIL 61

Db 370 NNSPFGACGTFRPAPMPPGPPPTTFSSLEFRQGVAKAESPAIILGLVAIL 429

Qy 62 VLILLMLV-----RRRLMPD-----CVGRGRLSPVDFL- 93

Db 430 LLILIALMLRLMRRLSKARRVLEELTVHLSVPDITLLINRRGPRPPRYDEPRP 489

Qy 94 AGDRPRAVPAVFMVLSLCLLPDDELPLFLTLASPSODGTEAPRGAW 145

Db 490 RGNPPPSAPC-----VPSGALLSNPVRLLATYAPRPGG---PPTPAW 534

RESULT 5

T36053 probable ABC-type transport protein cydCD - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C/Accession: T36053

R/Sanders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M. submitted to the EMBL Data Library, November 1998

A/Reference number: Z21595

A/Accession: T36053

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1172 <SAU>

A/Cross-references: EMBL:AL034355; PIDN:CAA22219.1; GSPDB:GN00070; SCODEB:SCDB.14

A/Experimental source: strain A3(2)

C/Genetics:

A/Gene: cydCD; SCODEB:SCDB.14

C/Superfamily: Mycobacterium tuberculosis probable ABC transporter RV0194; ATP-binding c

Query Match 7.8%; Score 83; DB 2; Length 1172;

Best Local Similarity 31.5%; Pred. No. 12; Matches 45; Conservative 12; Mismatches 68; Indels 18; Gaps 5;

Qy 55 LALSLIVLLMLVRRRLMPDPCVGRGRLSPVDFLAGDRPRAVPAVFMVLSLCL 114

Db 874 LAFEAVALGLPLAVOYROR-VRSARVYEVLDAPREVEPRQAPATPPVVEGLA 932

Qy 115 LLIP--DEDALPFLTL-----ASPSODGTEAPRGAWKILGFHYAALYPLA-- 161

Db 933 ARHPGQDRDALAGDLTLEQGRVAVVPGSGKTLT---AQVLRFLDSAGSYTLGV 989

Qy 162 -ACATAGTAHLGSLTSMHL 183

Db 990 DAVALAAGDVRRLVGLCAQDAHL 1012

RESULT 6

E83506 probable MFS transporter PA1108 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C/Accession: E83506

R/Stover, C.K.; Pham, X.Q.; Errin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bro

adman, S.; Yuan, Y.; Brody, L.V.; Coulter, S.N.; Folger, K.R.; Kab, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathor

A/Reference number: A82950; MUID:20437317; PMID:10984043

A/Accession: E83506

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1385 <STO>

A/Cross-references: GB:AE004541; GB:AE004091; NID:99947024; PIDN:AG04497.1; GSPDB:GN001

A/Experimental source: strain PA01

C/Genetics:

A/Gene: PA1108

Query Match 7.6%; Score 80.5; DB 2; Length 385;  
Best Local Similarity 29.0%; Pred. No. 6.1;  
Matches 54; Conservative 14; Mismatches 57; Indels 61; Gaps 13;

27 PGGKELQEGEVPSCHTSIPGLYHACIASLIVLLAMVRRRLQPCVGRP-- 84  
Db 148 PGGQWMLDLGHP-----AVFATLAVLASLALV---PMPPT---RL 189  
Qy 85 -GLPSPVDF-----LAGRP---RAVPAVFMVLLSLC---LLLPEDALP---LT 127  
Db 190 AGTPPATLAIIRRYLADRPLOTALVAVLAVFSPVAGPFWGDLPGIGFWIGLA 249  
Qy 128 LMSAPQDC--KTEAPRGAMKI-----IGLFHYALVYPLAACATAGTAAHLGST--- 177  
Db 250 IAIAGSLGALLRRRLPR-TMNSARVRLGL-----ALAAAGATTAQTLLAAVGVA 297  
Qy 178 --LSMA 181  
Db 298 EGLYMA 303

## RESULT 7

HNUTPH  
nucleocapsid protein N - Prospect Hill virus (strain PHV-1)  
N:Alternate names: nucleoprotein N  
C:Species: Prospect Hill virus  
C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 24-Jul-1997  
C:Accession: A34681  
R:Partridge, M.A.; Kang, C.Y.  
Virology 175, 167-175, 1990  
A:Title: Nucleotide sequence analysis of the S genomic segment of Prospect Hill virus:  
A:Reference number: A34681; MUID:90177212; PMID:2309440  
A:Accession: A34681  
A:Molecule type: genomic RNA  
A:Residues: 1-433 <PAR>  
C:Genetics:  
A:Gene: N  
A:Map position: segment S  
A:Superfamily: Hantaan virus nucleocapsid protein  
C:Keywords: nucleocapsid; nucleoprotein

Query Match 7.5%; Score 80; DB 1; Length 433;  
Best Local Similarity 24.1%; Pred. No. 7.7;  
Matches 35; Conservative 21; Mismatches 55; Indels 34; Gaps 6;

Qy 23 YIDPQGXELQEGEVPSCHTSIPGLYHACIASLIVLL---LAMVRRRLQMPD 78  
Db 178 YVSMPTQSTMKABELTPGRFRITVCGLPQAQIMANITSPMGVIGAFVVD---WAD 234  
Qy 79 CYRG-----RGLPS-PVDFLAGDRPRAVPAVFMVLLSLC---LPDE 120  
Db 235 KYKAFLODKCFLEAPRPQAGAEAFLLSTR-----ATLMKQAVLDITHLPDI 285  
Qy 121 DALPFLTLASAPSDGKTEAPRGAM 145  
Db 286 DALVELAASGDFLPLDLENPHAM 310

## RESULT 8

A49508  
protein-tyrosine kinase (EC 2.7.1.112) trkE precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 04-Feb-2000  
C:Accession: A49508; 138358; 537402  
R:Di Marco, E.; Cutuli, N.; Guerra, L.; Cancedda, R.; De Luca, M.  
J. Biol. Chem. 268, 24290-24295, 1993  
A:Title: Molecular cloning of trkE, a novel trk-related putative tyrosine kinase receptor  
A:Reference number: A49508; MUID:94043265; PMID:8226977  
A:Accession: A49508  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-876 <DIA>  
A:Cross-references: EMBL:X74979; NID:g400462; PIDN:CAA52915.1; PID:g400463

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; discoidin I amino-term  
C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase  
F:30-185/Domain: discoidin I amino-terminal homology <DNI>  
F:571-875/Domain: protein kinase homology <KN>  
F:579-587/Region: protein kinase ATP-binding motif

Query Match 7.5%; Score 80; DB 2; Length 876;  
Best Local Similarity 20.6%; Pred. No. 16;  
Matches 42; Conservative 26; Mismatches 74; Indels 62; Gaps 9;

Qy 7 NOTSPGATEDSYSGWYIDPQ---CGXELQEGEVPSCHT-SIPGLYHACIASLIL 61  
Db 370 NNSSPALGCTPPPAWMPGPPPTNSSLLEPRQGVAAAGSPTRILIGLVAIIL 429  
Qy 62 VLLILAMLV-----RRRLQMPD-----CYGRGLSPVDFLAGDR 97  
Db 430 LLLIATLMLRLHWRLLSKERRVLEBELTVHLVPGDTLLINRRPPRPPEPY--QEP 488  
Qy 98 PPAVPAVFMVLLSLC-----LLPDEDLPFLTLASAPSDGKTEAPRGAMKI 147  
Db 489 PRGNPPI-----SAPCVPGSAYSGDYMEPEKRAPLL---PPPPQN----- 527  
Qy 148 LGLFHYALVYPLAACATAGTAA 171  
Db 528 -SVPHVAAADIVTLQGVTOGNTYA 550

## RESULT 9

G02326  
transcription factor NFAT1 isoform B - human  
C:Species: Homo sapiens (man)  
C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 05-Nov-1999  
C:Accession: G02326  
R:Luo, C.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: H01056  
A:Accession: G02326  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-921 <LUD>  
A:Cross-references: EMBL:U43341; NID:G1353773; PIDN:AAC50866.1; PID:G1353774  
C:Genetics:  
A:Gene: NFAT1

Query Match 7.5%; Score 80; DB 2; Length 921;  
Best Local Similarity 23.5%; Pred. No. 17;  
Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;

Qy 3 QPAGNOTSPGATEDSYSGWYIDPQGX-----ELQEGEV-PSCHTSIPGLYHAC 54  
Db 240 RPASRSSSPGAKRRHSCAALVALPPGASPPQSRKSPSPSSHVAPODHS--PAG--YPP 296  
Qy 55 LASLIVLLILAMLVRR-----ROLW-----PDVYGRP---GLP---SPVPLA--- 94  
Db 297 VAGSAVIMALNSLATDPCGIPRGMKTPSPVSAASXKGLPRHLYPAVEFLGPCE 356  
Qy 95 -GDRPAVPAVFMVLLSLC---LLPDEDALPFLTL---AAP-----SQG----- 136  
Db 357 QGERNSAPESILVPPTPKPLV---AIPICSIPTVATLPLEPLSSQSGSYELRIE 413  
Qy 137 -----KTEAPRGAMKILGLFHYALVYPLAACATAGTAAHLGSLTSMANLGV 185  
Db 414 VQRPYHRAHYETESGRGVK-----APFGGHPVQVLHG--YVENKPLGL 456  
Qy 186 QVW 188  
Db 457 QIF 459

## RESULT 10

S47153  
type IV prepilin peptidase (EC 3.4.99.-) BfpP - Escherichia coli  
N:Contains: type IV pilin N-methyltransferase (EC 2.1.1.-)





A:Residues: 1-910 <RES>  
A:Cross-references: GB:I26525; NID:g432480; PIDN:AAA1089.1; PID:g432481  
C:Genetics:  
A:Gene: Ptk-3  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; discoidin I amino-ter  
C:Keywords: ATP  
F:31-186/Domain: discoidin I amino-terminal homology <DN1>  
F:605-909/Domain: protein kinase homology <KIN>  
F:613-621/Region: protein kinase ATP-binding motif

Query Match 7.4%; Score 79; DB 2; Length 910;  
Best Local Similarity 22.1%; Pred. No. 21;  
Matches 36; Conservative 23; Mismatches 70; Indels 34; Gaps 7;

Qy 13 ATEDSYSGSWYIDPO---GQXELQPEGEVPSCHT-SIPPGYHACLASLSTLVLLLA 67  
Db 373 SSDTPPPAPMPPPPPTNFSLELEPRGQOPVAKAGSPTALIGCLVAITLLILITA 432  
Qy 68 MIV-----RRQLWPD-----CYRGRPGLPSPVDFLAGDRPRAVA 103  
Db 433 LMLMLHMRRLSKARVLEDELTMLSVPGDTLLNNRPGREPPPY-QEPRPKTPT 491  
Qy 104 AVFMVLLSLCLLPDED-ALPPTLASAPSDGKTEAPRGAW 145  
Db 492 H8APCVPGSALLSNPAYRLLATYANPPRPG--PTPAW 531

## RESULT 14

H83588  
Probable MFS transporter PA0458 [Imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #ext\_change 31-Dec-2000  
C:Accession: H83588  
C:Stoever, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Laidig, K.; Lim,  
..; Lozy, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: H83588  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-477 <STO>  
A:Cross-references: GB:A8004483; GB:A8004091; NID:g9946313; PIDN:AAG03847.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0458

Query Match 7.4%; Score 78.5; DB 2; Length 477;  
Best Local Similarity 26.5%; Pred. No. 12;  
Matches 54; Conservative 16; Mismatches 63; Indels 71; Gaps 11;

Qy 19 YGSWY---IDEPG---GXELQPE-GEVPSCHTSPPGYHACLASLSI----- 60  
Db 164 YASWMIPLINPVGLCLVAMKMLPDRSPVSRPDSIGFLFGGSMVLISIALEGIG 223  
Qy 61 -----LVLLMLVRRQLMPDVR-GRPGLSPVDFLAGDRPRAVPAVMV-LL 110  
Db 224 ELHLSHRLRVLLIGGLVLTAYMLRALRIDKPLP-----PSLFKATFVNGIL 273  
Qy 111 SSLCLLPDEDALPPLT-----LASAPSDGKTEAPRG-----AMKILGFHYA 154  
Db 274 GNIFARL-GSGALPPLTPLLQVGLGYPPSTAGMTMPLALFAMVAKPMAKPLIDFFGYR 332  
Qy 155 ALYYPLAACATAGTAHLGLSTL 178  
Db 333 KL-----LVGNTL 340

RESULT 15  
JCS412  
epidermal growth factor receptor-binding protein GRB-7 - human  
C:Species: Homo sapiens (man)

C:Date: 10-Jun-1997 #sequence\_revision 18-Jul-1997 #ext\_change 21-Jul-2000  
C:Accession: JCS412  
R:Richt, T.; Sasaki, H.; Akiyama, N.; Ishizuka, T.; Sakamoto, H.; Aizawa, S.; Sugimura, T  
Biochem. Biophys. Res. Commun. 232, 5-9, 1997  
A:Title: Molecular cloning of human GRB-7 co-amplified with CAb1 and c-ERBB-2 in primary  
A:Reference number: JCS412; MUID:97236270; PMID:9125150  
A:Accession: JCS412  
A:Molecule type: mRNA  
A:Residues: 1-532 <KIS>  
A:Cross-references: DDBJ:D41772; NID:g601890; PIDN:BA07827.1; PID:g601891  
C:Comment: This protein contains a pleckstrin domain which mediates protein-protein inter  
C:Genetics:  
A:Gene: GDB:GRB7  
A:Cross-references: GDB:129754; OMIM:601522  
C:Superfamily: pleckstrin repeat homology  
F:231-336/Domain: pleckstrin #status predicted <PLE>  
F:432-532/Domain: SH2 #status predicted <SH2>

Query Match 7.4%; Score 78.5; DB 2; Length 532;  
Best Local Similarity 28.3%; Pred. No. 13;  
Matches 39; Conservative 12; Mismatches 62; Indels 25; Gaps 7;

Qy 75 LMPDCVGRPGLPSPVDF-LAGDRPRAVPAVMVLLSLCLLPDEDAL-----P 124  
Db 17 LMP-APGPPGTPRPDPFLPEVYKRSQP-----LLIPTGRKLREERRATSLPSTPVP 70  
Qy 125 FLTLASAPSDGKTEAPRGANKIL-----GLFHYAALYPLAAC-----ATAGTAHLGLS 176  
Db 71 PSELSPSPQSPFILGSPSARGLLPRDASPHVVKYSEDGACRYEVAAGATARRVCBM 130  
Qy 177 TLSMAH-LGVQYWGQRAEC 193  
Db 131 LVGRNALSDERMGLVEEC 148

Search completed: March 13, 2003, 11:30:58  
Job time : 15.3931 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:27:42 ; Search time 10.2538 Seconds  
(without alignments)  
571.023 Million cell updates/sec

Title: US-09-816-653A-2

Perfect score: 1061  
Sequence: 1 MSOPAGNQTSPGATEDYSYG.....MAHLGVQVWQRAECPOVKRI 199

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patente AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85.5	8.1	380	4	US-08-924-345-2
2	84	7.9	919	1	US-08-336-343A-2
3	83	7.8	913	1	US-08-445-640-4
4	83	7.8	913	3	US-08-170-558-4
5	83	7.8	913	3	US-08-447-314-4
6	83	7.8	913	3	US-08-445-461-4
7	80	7.5	433	3	US-09-106-075A-86
8	80	7.5	699	5	PCT-US94-07297-39
9	80	7.5	921	1	US-08-396-479B-2
10	80	7.4	921	1	US-08-818-823-2
11	79	7.4	505	1	US-08-221-750A-5
12	78.5	7.4	3519	4	US-09-428-517-4
13	76.5	7.2	875	1	US-08-785-241-5
14	75	7.1	331	2	US-08-985-217-7
15	73.5	6.9	378	1	US-09-482-180A-2
16	73.5	6.9	578	1	US-08-651-740-3
17	73.5	6.9	578	2	US-09-073-594-3
18	73.5	6.9	578	3	US-09-275-925-3
19	73.5	6.9	578	3	US-09-275-925-3
20	73.5	6.9	636	1	US-08-653-740-5
21	73.5	6.9	636	2	US-09-073-594-5
22	72	6.8	652	4	US-09-310-463-4
23	72	6.8	652	4	US-08-842-248A-4
24	72	6.8	694	4	US-09-440-345A-1
25	69.5	6.6	369	4	US-08-462-509B-4
26	69.5	6.6	369	5	PCT-US95-05616-4
27	69.5	6.6	372	4	US-08-462-509B-6

28	69.5	6.6	377	5	PCT-US95-05616-6	Sequence 6, Appl1
29	69.5	6.6	389	2	US-08-846-705-2	Sequence 2, Appl1
30	69.5	6.6	402	4	US-08-462-509B-2	Sequence 2, Appl1
31	69.5	6.6	402	5	PCT-US95-05616-2	Sequence 2, Appl1
32	69.5	6.6	425	3	US-08-846-704-2	Sequence 2, Appl1
33	69.5	6.6	650	4	US-09-310-463-2	Sequence 2, Appl1
34	69.5	6.6	1205	1	US-07-908-245-2	Sequence 2, Appl1
35	69.5	6.6	402	3	US-08-846-704-4	Sequence 2, Appl1
36	68.5	6.5	425	4	US-09-479-128-2	Sequence 4, Appl1
37	68.5	6.5	591	1	US-08-188-582-7	Sequence 7, Appl1
38	68	6.4	591	1	US-08-646-715-7	Sequence 7, Appl1
39	68	6.4	1064	3	US-08-726-214-8	Sequence 8, Appl1
40	67.5	6.4	373	4	US-09-254-077A-12	Sequence 12, Appl1
41	67.5	6.4	2864	4	US-08-469-260A-394	Sequence 394, App
42	67.5	6.3	368	5	PCT-US93-11703-24	Sequence 24, Appl
43	67	6.3	389	2	US-08-789-354-2	Sequence 2, Appl1
44	67	6.3	389	2	US-09-110-937-2	Sequence 2, Appl1
45	67	6.3	389	3	US-09-110-937-2	Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-08-924-345-2  
; Sequence 2, Application US/08924345  
; Patent No. 6224878  
; GENERAL INFORMATION:  
; APPLICANT: LEONG-TACK Patricia  
; APPLICANT: LEGASTELOIS Isabelle, Christine, Marie-Andree  
; APPLICANT: AUDONNET Jean-Christophe, Francis  
; APPLICANT: RIVIERE Michel, Emile, Albert  
; TITLE OF INVENTION: Mutants and vaccines of the Infectious  
; TITLE OF INVENTION: Bovine Rhinotracheitis virus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LARSON AND TAYLOR  
; STREET: 727 SOUTH TWENTY-THIRD STREET  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/924,345  
; FILING DATE: 04-SEP-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/199,172  
; FILING DATE: 09-AUG-1994  
; APPLICATION NUMBER: FR 92 07930  
; FILING DATE: 26-JUN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SARRO, THOMAS P.  
; REGISTRATION NUMBER: 19396  
; REFERENCE/DOCKET NUMBER: XI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 892-8428  
; TELEFAX: (703) 920-7200  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 380 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-924-345-2  
Query Match 8.1%, Score 85.5, DB 4, Length 380;  
Best Local Similarity 27.6%, Pred. No. 0.18;

Matches 37; Conservative 12; Mismatches 38; Indels 47; Gaps 6;

QY 4 PANGQTSFGATEDYSGSWYIDEPQ---GGXELQPEGEVPSCHT-SIPGLYHACLASLIL 59  
DB 256 PACPTASPAASPSHAFSA-----AAPAAQAQAGDTPA-----RRRQLASILVPLC 302  
QY 60 ILVLLIAM-----LVRRLQ-----WPCVGRGGLSPV 90  
DB 303 VLVLILALCALATYNCALRRRLPCSRVRYKPRTCACGSGTCAGRPGRGAPAPATV 362  
QY 91 DFLAGDRPAVPA 104  
DB 363 VAL-GSRPKAPPLA 375

## RESULT 2

US-08-336-343A-2  
Sequence 2, Application US/0836343A  
Patent No. 5677144  
GENERAL INFORMATION:  
APPLICANT: Ullrich, Axel  
TITLE OF INVENTION: CKK-2, A No. 5677144el Receptor Tyrosine Kinase  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/336,343A  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7663-065  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 865-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 919 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-336-343A-2

Query Match 7.9%; Score 84; DB 1; Length 919;

Best Local Similarity 22.7%; Pred. No. 0.9; Mismatches 73; Indels 40; Gaps 7;

Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;  
QY 7 NOTSPGATEDYSGSWYIDEPQ---GGXELQPEGEVPSCHT-SIPGLYHACLASLIL 61  
DB 370 NNSSPALGTFPPAPWMPGPPPTNFSLELEPRQCPVAKRGSPFALIGLVAILL 429  
QY 62 VLLILAMLV-----RRRLQMPD-----CYRGRGLSPVDFL----- 93  
DB 430 LLLIILALMLRMRLHMRLLSKARRVLEELTVHLISVPGDTLLNNRPGPREPPYQEP 489  
QY 94 AGDRPAVPAVAVFVLLSSCLLPDEDALPFLTLASPSODGTEAPRGAW 145  
DB 490 RGNPPHSAPC---VNGSALLSNPVRLLATYARPPRGPG---PPTPAW 534

RESULT 3  
US-08-445-640-4  
Sequence 4, Application US/08445640  
Patent No. 5709858  
GENERAL INFORMATION:

APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,640  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-445-640-4

Query Match 7.8%; Score 83; DB 1; Length 913;

Best Local Similarity 22.7%; Pred. No. 1.2; Mismatches 73; Indels 40; Gaps 7;

Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;  
QY 7 NOTSPGATEDYSGSWYIDEPQ---GGXELQPEGEVPSCHT-SIPGLYHACLASLIL 61  
DB 370 NNSSPALGTFPPAPWMPGPPPTNFSLELEPRQCPVAKRGSPFALIGLVAILL 429  
QY 62 VLLILAMLV-----RRRLQMPD-----CYRGRGLSPVDFL----- 93  
DB 430 LLLIILALMLRMRLHMRLLSKARRVLEELTVHLISVPGDTLLNNRPGPREPPYQEP 489  
QY 94 AGDRPAVPAVAVFVLLSSCLLPDEDALPFLTLASPSODGTEAPRGAW 145  
DB 490 RGNPPHSAPC---VNGSALLSNPVRLLATYARPPRGPG---PPTPAW 534

## RESULT 4

US-08-170-558-4  
Sequence 4, Application US/08170558  
Patent No. 6001621  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.

APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will P.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/170,558  
FILING DATE: 20-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1  
TELEPHONE: 415/952-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-170-558-4  
Query Match 7.8%; Score 83; DB 3; Length 913;  
Best Local Similarity 22.7%; Pred. No. 1.2; Mismatches 20; Indels 40; Gaps 7;  
Matches 39; Conservative 20; Mismatches 20; Indels 40; Gaps 7;  
Qy 7 NOTSPGATEDVSGSWYIDEPQ---GXELOPEGEVPSCH-TSIPGCLVHACLASLSIL 61  
Db 370 NNSSPALGCTFPAPWMPGPPPTNFSLSLEPRGQPPVAKPGSGPTALIGCTVAIILL 429  
Qy 62 VLLIAMLV-----RRQLMPD-----CYGRPGCLPSPVDL----- 93  
Db 430 LLLIILMLWRLLHWRLLSKAERVLSEELTVHLSVPGDTIILNNRPPRPPPYQERPP 489  
Qy 94 AGDRPRAVPAVFMVLTLSLCILLPDEDALEPFLTLASAPSDGKTEAPRGAW 145  
Db 490 RGNPPHSAPC---VPNGSALLLSNPAYRLLATYARPPRGPC---PPTPAW 534  
RESULT 5  
US-08-447-314-4  
Sequence 4, Application US/08447314  
Patent No. 6087144  
GENERAL INFORMATION:  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will P.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,314  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1D2  
TELEPHONE: 415/952-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-447-314-4  
Query Match 7.8%; Score 83; DB 3; Length 913;  
Best Local Similarity 22.7%; Pred. No. 1.2; Mismatches 20; Indels 40; Gaps 7;  
Matches 39; Conservative 20; Mismatches 20; Indels 40; Gaps 7;  
Qy 7 NOTSPGATEDVSGSWYIDEPQ---GXELOPEGEVPSCH-TSIPGCLVHACLASLSIL 61  
Db 370 NNSSPALGCTFPAPWMPGPPPTNFSLSLEPRGQPPVAKPGSGPTALIGCTVAIILL 429  
Qy 62 VLLIAMLV-----RRQLMPD-----CYGRPGCLPSPVDL----- 93  
Db 430 LLLIILMLWRLLHWRLLSKAERVLSEELTVHLSVPGDTIILNNRPPRPPPYQERPP 489  
Qy 94 AGDRPRAVPAVFMVLTLSLCILLPDEDALEPFLTLASAPSDGKTEAPRGAW 145  
Db 490 RGNPPHSAPC---VPNGSALLLSNPAYRLLATYARPPRGPC---PPTPAW 534  
RESULT 6  
US-08-445-461-4  
Sequence 4, Application US/08445461  
Patent No. 6096527  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,461

FILING DATE: 22-MAY-1995  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/170558  
 FILING DATE: 20-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/157563  
 FILING DATE: 23-NOV-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hasak, Janet E.  
 REGISTRATION NUMBER: 28,616  
 REFERENCE/DOCKET NUMBER: 854C3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-1896  
 TELEFAX: 415/952-9881  
 TELE: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 913 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-445-461-4

Query Match 7.8%; Score 83; DB 3; Length 913;  
 Best Local Similarity 22.7%; Pred. No. 1.2; Mismatches 73; Indels 40; Gaps 7;

Matches 39; Conservative 20; Mismatches 73; Indels 40; Gaps 7;  
 Oy 7 NOTSGATEDSYGSWYIDEPQ---GKELQEGEVPSCH-TSIPGLYHACLASLSTL 61  
 Db 370 NNSBALGCTPPAPMPPGPPPTNSSLEPRGQPVAKPESPTAILGCLVAILL 429  
 Oy 62 VLLLLMLV-----RRQLMPD-----CVRGRPLSPVDFL----- 93  
 Db 430 LLLIALLMLHMRRLSKERRVLEELTVLHVPDITLNNRPPRRPPYQPRP 489  
 Oy 94 AGDPRPAVPAVFWLSSCLLPDPEDALPFLTLASAPSDGKTEAPRGAW 145  
 Db 490 RGNPPHAPC---VPMGSLLSNPAYRLLATYARPRGPG--PPTPAW 534

RESULT 7  
 US-09-106-075A-86  
 Sequence 86, Application US/09106075A  
 Patent No. 6316250  
 GENERAL INFORMATION:  
 APPLICANT: Hjelte, MD, Brian  
 TITLE OF INVENTION: Molecular Clones Producing Recombinant DNA Antigens of  
 FILE REFERENCE: 10312-8U1, Hjelte et al. (210312.0009)  
 CURRENT FILING DATE: US/09/106,075A  
 PRIOR APPLICATION NUMBER: 08/210,762  
 PRIOR FILING DATE: 1994-03-22  
 PRIOR APPLICATION NUMBER: 08/144,035  
 PRIOR FILING DATE: 1993-10-26  
 PRIOR APPLICATION NUMBER: 08/120,096  
 PRIOR FILING DATE: 1993-09-13  
 PRIOR APPLICATION NUMBER: 08/111,519  
 PRIOR FILING DATE: 1993-08-25  
 NUMBER OF SEQ ID NOS: 90  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 86  
 LENGTH: 433  
 TYPE: PRT  
 ORGANISM: Prospect Hill virus  
 US-09-106-075A-86

Query Match 7.5%; Score 80; DB 4; Length 433;  
 Best Local Similarity 24.1%; Pred. No. 0.91; Mismatches 55; Indels 34; Gaps 6;  
 Matches 35; Conservative 21; Mismatches 55; Indels 34; Gaps 6;  
 Oy 23 YIDPQGGKELQEGEVPSCHTSIPGLYHACLASLSTLVL---LMLVRRQLMPD 78

Db 178 YVSMPTASTMKAELLTPGFRITVGLFPQAIDMRNLISSVMGVIAFAFVND---MAD 234  
 Oy 79 CVRG-----RCGLPS-PYDPLAGRPRAVPAVFWLSSCLL---LPDE 120  
 Db 235 KVNAPLDQCKCPFKAPRRPGQPGAGNEFLSIR-----AYLMNRQAVLDLTHLPDI 285  
 Oy 121 DALPFLTLASAPSDGKTEAPRGAW 145  
 Db 286 DALVELAASGDPTLPDLENPHAAW 310

PCT-US94-07297-39  
 Sequence 39, Application PC/TUS9407297  
 GENERAL INFORMATION:  
 APPLICANT: Arai, Naoko  
 APPLICANT: Masuda, Esteban S.  
 APPLICANT: Tokumitsu, Hiroshi  
 TITLE OF INVENTION: PURIFIED COMPONENTS OF MAMMALIAN  
 TITLE OF INVENTION: TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS  
 NUMBER OF SEQUENCES: 41  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: John H. C. Blasdale, Schering-Plough Corporation, W-3-W  
 STREET: One Giralda Farms  
 CITY: Madison  
 STATE: New Jersey  
 COUNTRY: USA  
 ZIP: 07940-1000  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: Apple Macintosh IIcx  
 SOFTWARE: Microsoft Word 5.1a  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/07297  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/222,626  
 FILING DATE: 04-APR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/148,061  
 FILING DATE: 05-NOV-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/113,971  
 FILING DATE: 30-AUG-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/099,998  
 FILING DATE: 30-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/088,483  
 FILING DATE: 06-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Blasdale, John H. C.  
 REGISTRATION NUMBER: 31,895  
 REFERENCE/DOCKET NUMBER: DX0392K4  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-822-7398  
 TELEFAX: 201-822-7039  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 699 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US94-07297-39

Query Match 7.5%; Score 80; DB 5; Length 699;  
 Best Local Similarity 23.5%; Pred. No. 1.7; Mismatches 76; Indels 80; Gaps 15;  
 Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;  
 Oy 3 QPAGNOTSGATEDSYGSWYIDEPQGX-----ELQEGEV-PSCHTSIPGLYHAC 54

Db 21 RPARSSSSPGAKRRHSCAEALVALPPGASPORSRSPSPQPSHVAPODHG--PAG--YPP 77  
Qy 55 LASLSTIVLLMLVRR-----ROLW-----PDCVGRP---GLP-----SPVDFLA--- 94  
Db 78 VAGSAVIMDLNLSLTDSCGIPPRMMKTSPPSPSVSAAPSAGLPRHIYPAVEFLGPCE 137  
Qy 95 -GDRPRAVPAVFMVLLSLCLLPDEDALPFLTL--ASAP-----SQDG----- 136  
Db 138 QGERRNAPESILLVPTPKPLVP---AIPICSIPTVASLPLEWPLSSQSGSYELRIE 194  
Qy 137 -----KTEAPRGAMKILGLFHYALYYPLAACATAGHTAAHLGSTLSMAHLGV 185  
Db 195 VOPKPHRAHYETEGSRGAVK-----APTGHFVVQLHG--YMENKPLGL 237  
Qy 186 QVM 188  
Db 238 QIF 240

## RESULT 9

US-08-396-479B-2  
Sequence 2, Application US/08396479B  
Patent No. 5612455  
GENERAL INFORMATION:  
APPLICANT: HOBY, Timothy  
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/396,479B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osmann, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59450-1/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 210 277299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 921 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-396-479B-2

Query Match 7.5%; Score 80; DB 1; Length 921;  
Best Local Similarity 23.5%; Pred. No. 2.5;  
Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;

Qy 3 QPANGOTSPGATEDYISGSWYIDEPQGX-----ELOPEGEV--PSCHTSTPPGLYHAC 54  
Db 240 RPARSSSSPGAKRRHSCAEALVALPPGASPORSRSPSPQPSHVAPODHG--PAG--YPP 296  
Qy 55 LASLSTIVLLMLVRR-----ROLW-----PDCVGRP---GLP-----SPVDFLA--- 94  
Db 297 VAGSAVIMDLNLSLTDSCGIPPRMMKTSPPSPSVSAAPSAGLPRHIYPAVEFLGPCE 356  
Qy 95 -GDRPRAVPAVFMVLLSLCLLPDEDALPFLTL--ASAP-----SQDG----- 136

Db 357 QGERRNAPESILLVPTPKPLVP---AIPICSIPTVASLPLEWPLSSQSGSYELRIE 413  
Qy 137 -----KTEAPRGAMKILGLFHYALYYPLAACATAGHTAAHLGSTLSMAHLGV 185  
Db 414 VOPKPHRAHYETEGSRGAVK-----APTGHFVVQLHG--YMENKPLGL 456  
Qy 186 QVM 188  
Db 457 QIF 459

## RESULT 10

US-08-818-823-2  
Sequence 2, Application US/08818823  
Patent No. 5708158  
GENERAL INFORMATION:  
APPLICANT: HOBY, Timothy  
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,823  
FILING DATE: 14-MAR-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/396,479  
FILING DATE: 02-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Osmann, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59450-1/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 210 277299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 921 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-818-823-2

Query Match 7.5%; Score 80; DB 1; Length 921;  
Best Local Similarity 23.5%; Pred. No. 2.5;  
Matches 57; Conservative 30; Mismatches 76; Indels 80; Gaps 15;

Qy 3 QPANGOTSPGATEDYISGSWYIDEPQGX-----ELOPEGEV--PSCHTSTPPGLYHAC 54  
Db 240 RPARSSSSPGAKRRHSCAEALVALPPGASPORSRSPSPQPSHVAPODHG--PAG--YPP 296  
Qy 55 LASLSTIVLLMLVRR-----ROLW-----PDCVGRP---GLP-----SPVDFLA--- 94  
Db 297 VAGSAVIMDLNLSLTDSCGIPPRMMKTSPPSPSVSAAPSAGLPRHIYPAVEFLGPCE 356  
Qy 95 -GDRPRAVPAVFMVLLSLCLLPDEDALPFLTL--ASAP-----SQDG----- 136  
Db 357 QGERRNAPESILLVPTPKPLVP---AIPICSIPTVASLPLEWPLSSQSGSYELRIE 413  
Qy 137 -----KTEAPRGAMKILGLFHYALYYPLAACATAGHTAAHLGSTLSMAHLGV 185

Db 414 VQPKPHRAHYETSGRAVK-----APTGHFVVLQHC-YMENKPLGL 456  
QY 186 QVW 188  
Db 457 QIF 459

RESULT 11  
US-08-221-750A-5  
Sequence 5, Application US/08221750A  
Patent No. 5643747  
GENERAL INFORMATION:  
APPLICANT: Baker, Steven M.  
APPLICANT: Deitch, Robert A.  
TITLE OF INVENTION: Genes for the Export of Pertussis  
TITLE OF INVENTION: Holotoxin  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millitt Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/221,750A  
FILING DATE: 31-MAR-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/031,619  
FILING DATE: 15-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Alice O.  
REGISTRATION NUMBER: 33,542  
REFERENCE/DOCKET NUMBER: ACC93-01A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 505 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-221-750A-5

Query Match 7.4%; Score 79; DB 1; Length 505;  
Best Local Similarity 26.3%; Pred. No. 1.5;  
Matches 54; Conservative 18; Mismatches 71; Indels 62; Gaps 11;  
QY 5 AGNQTSPATDYSYGVSWYIDEPGCGXELPREGVPSCHTS--IPGLYHACLASISL 59  
Db 220 AANDANPSSAVNMLICAMIVASAGV-----LCLASLILVPGIYITLLISLOPL 270  
QY 60 ILVLLMLMLVRRQLWPDVC-----RGRGLSPSPVDPLAGDRPRAVPAV----- 105  
Db 271 FLVLLLPALGRMTNAMLGALVRALVPMALGTAVGLSDVLAGALPALPORAFTBLR 330  
QY 106 FVVLISLSC-----LLLPDEBALPFLTLASA-----PSODGKTEARPGAWKILG 149  
Db 331 STMALATTCATATML-----LTLVPLASSVNAAGLRRRLPRAAHGLAQAHRQA----- 380  
QY 150 LFHYAALY---PLAACATAG-ATA 170  
Db 381 -----AARQYARRPAAAAAAGPHQA 401  
RESULT 12

US-09-428-517-4  
Sequence 4, Application US/09428517  
Patent No. 6251636  
GENERAL INFORMATION:  
APPLICANT: Betlach, Mary C.  
APPLICANT: Shah, Sanjay Krishnakant  
APPLICANT: McDaniel, Robert  
APPLICANT: Tang, Li  
TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 30062-20029.00  
CURRENT APPLICATION NUMBER: US/09/428,517  
CURRENT FILING DATE: 1999-10-28  
EARLIER APPLICATION NUMBER: 60/120,254  
EARLIER FILING DATE: 1999-02-16  
EARLIER APPLICATION NUMBER: 60/106,100  
EARLIER FILING DATE: 1998-10-29  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 4  
LENGTH: 3519  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Recombinant  
OTHER INFORMATION: Oleandolide PKS  
US-09-428-517-4

Query Match 7.4%; Score 78.5; DB 4; Length 3519;  
Best Local Similarity 23.5%; Pred. No. 23;  
Matches 54; Conservative 25; Mismatches 70; Indels 81; Gaps 12;  
QY 11 PGATEDYSYGVSWYIDEPGCGXELPREGVPSCHTSIPGLYHACLASISLVLMLV 70  
Db 2642 PALDEVLGGGWLFFVPRG-----LADGVARVAVAVTA 2676  
QY 71 R-----RQLMPDCVGRGLSPVDPLAGDRP-----APA--AVPMVL 109  
Db 2677 RGEVSVELDPTPRPRRAYAEVAGR-GVSGVSLSDRRRSHRPVPAGLASLVL 2735  
QY 110 LLSLSC-----LLLPDEBALPFLTLASAPSDGKTEAPRGA--W--KIIGLFH-- 152  
Db 2736 AQAALVDLGRVGEPRMLVLVTRDA-----VVAQSPDAGAVIDPVOAOWMGFGFVGLSHPE 2790  
QY 153 --YAALYPLAA---CATAGHPRAHILGSTLSMAHL-----GVQVWOR 190  
Db 2791 LMGLIDLPEAPRPGSTCDHYADLLATVVASAGFEDQAVAVRSGVAVR 2840

RESULT 13  
US-08-785-241-5  
Sequence 5, Application US/08785241  
Patent No. 5695963  
GENERAL INFORMATION:  
APPLICANT: McKnight, Steven L.  
APPLICANT: Russell, David W.  
APPLICANT: Tian, Hui  
TITLE OF INVENTION: Endothelial PAS Domain Protein  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,241  
FILING DATE: 17-JAN-1997



CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A.  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: US-09-816-653a-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 875 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-816-653a-2

Query Match  
Best Local Similarity 7.2%; Score 76.5; DB 1; Length 875;  
Matches 61; Conservative 17; Mismatches 93; Indels 95; Gaps 14;

QY 10 SPG-ATEDYSYGSWYIDEPQG-GXELQPEG-VP-----40  
DB 407 TPEDALISLDFSGQNFDEPSAYGKALILPPQPMVSGLRSHSAQESGLPAFTVPADTP 466  
QY 41 -----SCHTSIPPGLYHACLAS-LSLVL-LTLLMLVRR-----73  
DB 467 GNTTPASSSSSCSTPSSSEEDYSSLENPLKIEVLEKLFMDTEPRDPGSGTQDSELD 526  
QY 74 -----QLMPDCVGRGRLSPVDPLAGDRPRAVPAVWVLLSSLCLELD 119  
DB 527 ETLAPYIPMDGDFOLSPIC-----PEEPLMPSPPOPTPHCFSTSTISFOPLTG 577  
QY 120 EDALPPLTASAPSO-DKTEAPGAWKILGLFHYALYPLAACATAGTAHLG-- 175  
DB 578 ATHGPPF-LDKYPPQLSKRKTESH-WPMNSIFPDAGSGSLPCCGASSTLSMGGR 634  
QY 176 STLSW-----AHLGVQW-----QRAE 192  
DB 635 SNTQWPPDPLHFGPTKMPVGDOSAE 660

RESULT 14  
US-09-816-653a-2  
Sequence 7, Application US/08986217  
Patent No. 5914249  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Shah, Purvi  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: CELL-CYCLE PHOSPHOPROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESS: Inocyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/986,217  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PR-0452 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 331 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 2072296  
US-09-816-653a-2

Query Match  
Best Local Similarity 7.1%; Score 75; DB 2; Length 331;  
Matches 49; Conservative 25; Mismatches 84; Indels 58; Gaps 9;

QY 11 PGATEDYSYGSWYIDEPQGXELQPE--GEVPSCHTSIPPGLYHACLASLVLTLA 67  
DB 23 PMGAEPMALGSPSPKPSAGQFLDPLGDIPTVTPOP-----RPSLGIMEV-----71  
QY 68 MLVRRQLMPDCVGRPGPS-----PV-----GDRPRAV-PAAVWVLL 110  
DB 72 -----RPLSGSGSPQPVLPHTIKDKSGAPVRSIYDDVASGLGSTRNTRKXASVLAH 127  
QY 111 SSVCLLIPEDAL-----PPLTASAPSDGKTEAPGAWKILGL 150  
DB 128 TPLSGAIPSPAVFSPATIGSRKTTLSPAQMDPFYTGDAITSDOLD--DTWTVRG 184  
QY 151 FHYALYPLAACATAGTAHLGSLTSMHLGQ 186  
DB 185 PQASASYILLQFAQYGNITKIVMNNNNHHIQYQ 220

RESULT 15  
US-09-816-653a-2  
Sequence 2, Application US/09482180A  
Patent No. 6361985  
GENERAL INFORMATION:  
APPLICANT: Conklin, Darrell C.  
APPLICANT: Yamamoto, Gayle  
APPLICANT: Gao, Zeren  
APPLICANT: Whitmore, Theodore E.  
APPLICANT: Jasper, Stephen  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING  
FILE REFERENCE: 98-80  
CURRENT APPLICATION NUMBER: US/09/482,180A  
CURRENT FILING DATE: 2000-01-12  
PRIOR APPLICATION NUMBER: 60/115,721  
PRIOR FILING DATE: 1999-01-12  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 2  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-816-653a-2

Query Match  
Best Local Similarity 6.9%; Score 73.5; DB 4; Length 378;  
Matches 53; Conservative 16; Mismatches 82; Indels 61; Gaps 11;

QY 29 GGELOPEGEVVSCHTSIPPGLYHAC-LASLVLTLMLVRRQLMPDCVGRPGLP 87  
DB 15 GRSGGLPKG-----PMLCRICWLVSYSLVLLGLFLRKAAK--AGDPTAH 62  
QY 88 SPVDPLAGDRP-----AVPAVWVLLSLC-----LLIP--DEDALP 124  
DB 63 QP--FWAPPTPRHSRCPNHTVSSASLSLPSRRLFTYHCHNPFSLLEPSGCSKDTPL 120

Qy 125 FLTLAGPSQDGTENRGAWKILGLF---HYAALYFPLAACATAGHTAHLIG----- 175  
Db 121 LLAIKSQPGHVERRAAIRSTWGRVGGWARGROLKIVFLLGVAGSA--PPAQLLAYESREF 178  
Qy 176 -STLSW-----AHLGVQVWQRAECPO 195  
Db 179 DDILQWDFTEDPFNLTLELHLORWVVAACPQ 210

Search completed: March 13, 2003, 11:31:33  
Job time : 14.2538 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:25:47 ; Search time 25.0649 Seconds  
(without alignments)  
1057.929 Million cell updates/sec

Title: US-09-816-653A-2  
Perfect score: 1061  
Sequence: 1 MSOPAGNQTSFGATEDYSYG.....MAHLGVQVWRAECPOVPEKI 199

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 segs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq.101002.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1048	98.8	487	23	AAO17374 Human retinoic acid
2	1048	98.8	667	22	AAU04557 Human Str66 homolo
3	1048	98.8	667	22	AAAB8572 Human hydrophobic
4	1048	98.8	667	23	AAU78575 Human Str66 (PRO10
5	1048	98.8	667	23	AAU83631 Human PRO protein,
6	1047	98.7	667	23	AAU80548 Human NOV6c protei
7	1043	98.3	662	23	AAU80547 Human NOV2b protei
8	1020	96.1	689	23	AAU91278 Human NOV2b protei
9	986.5	93.0	658	22	AAU04558 Human Str66 homolo
10	986.5	93.0	658	23	AAU78576 Human Str66 protei

11	985.5	92.9	653	23	ABB05046 Human NOV2a protei
12	898	84.6	660	23	AAU91277 Human NOV2a protei
13	488	46.0	550	22	AAU94108 Human protein segn
14	87.5	8.2	1070	22	ABG02155 Novel human diagno
15	85.5	8.1	380	15	AAU48063 Sequence of polype
16	84	7.9	563	18	AAW34674 Human mammary carc
17	84	7.9	624	21	AAU54286 Human pancreatic c
18	84	7.9	624	22	AAU73167 Human mammary carc
19	84	7.9	624	16	AAU75502 Human mammary carc
20	84	7.9	919	16	AAU75504 Human mammary carc
21	84	7.9	919	18	AAU34672 Human mammary carc
22	80.5	7.6	130	20	AAU34629 Human secreted pro
23	80	7.5	433	18	AAU1875 Human mammary carc
24	80	7.5	876	18	AAU34675 Human mammary carc
25	80	7.5	876	23	AAU10542 Human epithelial d
26	80	7.5	882	18	AAU34673 Human mammary carc
27	80	7.5	921	17	AAU02248 Human transcritp
28	80	7.5	1171	19	AAU85037 Green fluorescent
29	80	7.5	1181	19	AAU85038 NPAT-green fluore
30	79.5	7.5	485	22	ABU19631 Novel human diagno
31	79	7.4	505	18	AAU23675 Bordetella pertuss
32	79	7.4	699	16	AAU66875 Human NF-A120, p
33	78.5	7.4	3519	21	AAU92709 S. antibiolicus 8,
34	78	7.4	263	22	AAU50117 Propionibacterium
35	78	7.4	474	22	ABG09574 Novel human diagno
36	78	7.4	501	22	ABU13111 Novel human diagno
37	77.5	7.3	914	16	AAU71100 Novel human diagno
38	77	7.3	535	22	ABG03456 Human polypeptide
39	76.5	7.2	215	22	AAU40761 Human polypeptide
40	76.5	7.2	337	22	ABU15237 Murine endothelial
41	76.5	7.2	875	19	AAU37098 Human protein SEQ
42	76	7.2	561	22	AAU45301 Human 51-binding a
43	76	7.2	2189	23	AAU79017 Human 51-binding a
44	75.5	7.1	466	23	AAU48999 Novel human diagno
45	75.5	7.1	740	22	ABG03831

#### ALIGNMENTS

RESULT 1  
ID AAO17374 standard; Protein: 487 AA.  
XX AAO17374;  
AC  
XX  
XX  
08-JUL-2002 (first entry)  
DT  
XX  
XX Human retinoic acid responsive protein 53-57.  
DE  
XX Human: retinoic acid responsive protein 53.57; arrhythmia;  
KW bronchial asthma; peptic ulcer; diabetes; cancer; cytostatic; antitumor;  
KW antidiabetic; antidiabetic; endocrine; cardiac;  
KW embryonic development disorder; gene therapy.  
OS  
XX Homo sapiens.  
XX  
XX WO200226808-A1.  
XX  
XX 04-APR-2002.  
XX  
XX 20-AUG-2001; 2001WO-CN01252.  
XX  
XX 23-AUG-2000; 2000CN-0119715.  
XX  
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
XX  
XX Mao Y, Xie Y;  
XX WPI, 2002-330093/36.  
XX N-PSDB; AAU45968.  
XX  
XX Human retinoic acid responsive protein 53.57 and encoding

[illegible]

FT	Region	317..323	/label=N_myristoylation_site
FT	Region	341..347	/label=N_myristoylation_site
FT	Region	364..375	/note="Prokaryotic membrane lipoprotein attachment site"
FT	Domain	364..380	/label=Transmembrane_domain
FT	Domain	431..451	/label=Transmembrane_domain
FT	Domain	474..489	/label=Transmembrane_domain
FT	Region	525..531	/label=N_myristoylation_site
FT	Region	550..535	/label=Transmembrane_domain
FT	Region	627..633	/label=N_myristoylation_site
FT	Region	631..637	/label=N_myristoylation_site
FT	Region	640..646	/label=N_myristoylation_site
FT	Region	661..667	/label=N_myristoylation_site
XX			
PN	WO200151635-A2.		
PD	19-JUL-2001.		
XX			
PF	11-JAN-2001; 2001MO-US00847.		
PR	13-JAN-2000; 2000US-0175849.		
PR	14-APR-2000; 2000US-0197089.		
PR	29-AUG-2000; 2000US-0228914.		
PA	(GETH ) GENENTECH INC.		
XX			
PI	Pennica D, Smith V, Wood WI;		
DR	WPI; 2001-442146/47.		
XX			
DR	N-PSDB; AAS08630.		
XX			
PT	Human nucleic acid encoding a PRO10282 or PRO19576 polypeptide (the PRO		
PT	polypeptides are referred to as Str66 polypeptides), useful in		
PT	molecular biology, including use as hybridisation probes, and in		
PT	chromosome and gene mapping -		
PS	Claim 24; Fig 2; 15pp; English.		
XX			
CC	The sequence is human PRO10282, a homologue of murine Str66, a		
CC	retinoic acid responsive protein thought to play an important role in		
CC	early dorboventral limb patterning during development and later in the		
CC	control of endochondral ossification. The gene for the Str66 homologue		
CC	is located on chromosome 15q23. The Str66 polypeptides, agonists,		
CC	antagonists or anti-Str66 antibodies are useful for preparation of a		
CC	medicament useful in the treatment of a condition which is responsive to		
CC	the Str66 polypeptides, agonists, antagonists or anti-Str66 antibodies.		
CC	The Str66 polypeptides may also be employed as molecular weight markers		
CC	for protein electrophoresis. The Str66 nucleic acids have applications in		
CC	molecular biology, including use as hybridisation probes, and in		
CC	chromosome and gene mapping. The antibodies and other anti-tumour		
CC	compounds may be used to treat various conditions, including those		
CC	characterised by overexpression and/or activation of amplified genes.		
CC	Exemplary conditions or disorders include benign or malignant tumours		
CC	(e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,		
CC	colorectal, prostate, pancreatic, lung, uterine, thyroid, hepatic		
CC	carcinomas, sarcomas, glioblastomas, and various head and neck tumours,		
CC	leukemias, and lymphoid malignancies), neuronal, glial, astrocytic,		
CC	hypothalamic and other glandular, macrophagal, epithelial, stromal and		
CC	biocellular disorders, inflammatory, angiogenic and immunogenic		
CC	disorders.		
XX			

SQ Sequence 667 AA:  
 Query Match 98.8%; Score 1048; DB 22; Length 667;  
 Best Local Similarity 99.0%; Pred. No. 4.3e-106;  
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 3 SOPAGNQTSPGATGTEYSGSWYIDEPQGGELQPEGEVPSCHTSIPGLYHACLASLSTL 61  
 QY 2 SOPAGNQTSPGATGTEYSGSWYIDEPQGGELQPEGEVPSCHTSIPGLYHACLASLSTL 61  
 Db 3 SOPAGNQTSPGATGTEYSGSWYIDEPQGGELQPEGEVPSCHTSIPGLYHACLASLSTL 62  
 QY 62 VLLLLMLVRRRQLMPDCVGRPGLPSPVDFIAGDRRAVPAVFWVLSSLCILLPPDD 121  
 Db 63 VLLLLMLVRRRQLMPDCVGRPGLPSPVDFIAGDRRAVPAVFWVLSSLCILLPPDD 122  
 QY 122 ALPFLTLAASPQDGKTEAPRGAMKILGLFYAALYYPLAACATAGHTAHLIGSTLSMA 181  
 Db 123 ALPFLTLAASPQDGKTEAPRGAMKILGLFYAALYYPLAACATAGHTAHLIGSTLSMA 182  
 QY 182 HLGQVQWQRAECPOVPKI 199  
 Db 183 HLGQVQWQRAECPOVPKI 200

RESULT 3  
 AAB8572  
 ID AAB8572 standard; Protein; 667 AA.  
 XX  
 AC AAB8572;  
 DT 04-JUN-2001 (first entry)  
 DE Human hydrophobic domain containing protein clone HP10713 #36.  
 XX  
 KW Human; hydrophobic domain; immunosuppressant; anti-HIV, neuroprotective;  
 KM antitumoric; vulnery; antitumor; osteoporosis; anti-inflammatory;  
 KM cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;  
 KW HIV infection; anaemia; burn; ulcer; osteoporosis; tumour; wound healing;  
 KM inflammatory bowel disease; nutritional supplement; appetite; vaccine;  
 KW behavioural characteristic; immune response.  
 XX  
 OS Homo sapiens.  
 PN MO200112660-AZ.  
 PD 22-FEB-2001.  
 PF 10-AUG-2000; 2000WO-JP05356.  
 XX  
 PR 17-AUG-1999; 99JP-0230344.  
 PR 07-SEP-1999; 99JP-0252551.  
 PR 01-OCT-1999; 99JP-0281132.  
 PR 22-OCT-1999; 98JP-0301624.  
 PR 04-NOV-1999; 99JP-0313877.  
 XX  
 PA (SAGA) SAGAMI CHEM RES CENT.  
 PA (PROT-) PROTEGENE INC.  
 PI Kato S, Kimura T;  
 DR MPI: 2001-160059/16.  
 DR N-PSDB: AAF94442.  
 XX  
 PT Human proteins with hydrophobic domains and the DNA which encode them  
 PT are useful for treating autoimmune disorders, burns and tumors and for  
 PT screening novel pharmaceuticals -  
 XX  
 PS Claim 1: Page 288-291; 518pp; English.  
 CC AAF94417 to AAF94516 encode the human proteins given in AAB8572 to  
 CC AAB88606 (I) which have a hydrophobic domain. (I) have immunosuppressant,  
 CC anti-HIV, neuroprotective, antitumoric, vulnery, antitumor,  
 CC osteoporosis, anti-inflammatory and cytostatic activities, and can be  
 CC used in gene therapy. (I) can be used as pharmaceuticals and as antigens

CC to prepare antibodies. DNA and cDNA (II) encoding (I) can be used as  
 CC probes for genetic diagnosis and gene sources for gene therapy or for  
 CC producing (I) in large quantities. Cells containing (II) are used for  
 CC the detection of ligands or receptors corresponding to membrane or  
 CC secretory proteins and to screen small molecule novel pharmaceuticals.  
 CC Antibodies directed to (I) can be used for the detection, quantification  
 CC and purification of (I). Activities of (I) may include cytokine and cell  
 CC proliferation/differentiation function, immune stimulating or suppressing  
 CC activity, haemostasis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory  
 CC activity. (I) and (II) can be used to treat autoimmune disorders e.g.  
 CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,  
 CC inflammatory bowel disease and tumours. (I) and (II) can also be used for  
 CC wound healing, as nutritional sources or supplements e.g. as amino acid,  
 CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,  
 CC processing and utilisation of dietary fat, protein, carbohydrate, vitamin,  
 CC vitamins and minerals, to effect behavioural characteristics, to affect  
 CC appetite, and can act as antigens in vaccines to raise an immune response  
 CC to the protein or another material cross-reactive with the protein.  
 CC  
 SQ Sequence 667 AA:

Query Match 98.8%; Score 1048; DB 22; Length 667;  
 Best Local Similarity 99.0%; Pred. No. 4.3e-106;  
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Db 2 SOPAGNQTSPGATGTEYSGSWYIDEPQGGELQPEGEVPSCHTSIPGLYHACLASLSTL 61  
 Db 3 SOPAGNQTSPGATGTEYSGSWYIDEPQGGELQPEGEVPSCHTSIPGLYHACLASLSTL 62  
 QY 62 VLLLLMLVRRRQLMPDCVGRPGLPSPVDFIAGDRRAVPAVFWVLSSLCILLPPDD 121  
 Db 63 VLLLLMLVRRRQLMPDCVGRPGLPSPVDFIAGDRRAVPAVFWVLSSLCILLPPDD 122  
 QY 122 ALPFLTLAASPQDGKTEAPRGAMKILGLFYAALYYPLAACATAGHTAHLIGSTLSMA 181  
 Db 123 ALPFLTLAASPQDGKTEAPRGAMKILGLFYAALYYPLAACATAGHTAHLIGSTLSMA 182  
 QY 182 HLGQVQWQRAECPOVPKI 199  
 Db 183 HLGQVQWQRAECPOVPKI 200

RESULT 4  
 AAU78575  
 ID AAU78575 standard; Protein; 667 AA.  
 XX  
 AC AAU78575;  
 DT 18-JUN-2002 (first entry)  
 DE Human StrA6 (PRO10282) protein.  
 XX  
 KW Human; cancer; StrA6; cytostatic; stromal disorder;  
 KW tumour; retinoid; Wnt; tumour antigen; leukaemia; lymphoid malignancy;  
 KW neuronal disorder; glial disorder; astrocytal disorder;  
 KW hypothalamic disorder; glandular disorder; macrophagal disorder;  
 KW epithelial disorder; blastocoele disorder; chromosome 15q23.  
 XX  
 OS Homo sapiens.  
 FH  
 FH Key Location/Qualifiers  
 FT 8..12  
 FT /note= "N Glycosylation site"  
 FT 50..56  
 FT /note= "N-myristoylation site"  
 FT 54..69  
 FT /note= "Transmembrane domain"  
 FT 102..119  
 FT /note= "Transmembrane domain"  
 FT 132..140  
 FT /note= "ATP/GTP binding site motif A"

PT Domain 148..166 /note= "Transmembrane domain"  
 PT Region 176..182 /note= "N-myristoylation site"  
 PT Domain 207..222 /note= "Transmembrane domain"  
 PT Region 241..247 /note= "N-myristoylation site"  
 PT Domain 301..320 /note= "Transmembrane domain"  
 PT Region 317..323 /note= "N-myristoylation site"  
 PT Region 341..347 /note= "N-myristoylation site"  
 PT Domain 364..380 /note= "Transmembrane domain"  
 PT Region 364..375 /note= "lipoprotein lipid attachment site"  
 PT Domain 431..451 /note= "Transmembrane domain"  
 PT Domain 474..489 /note= "Transmembrane domain"  
 PT Region 525..531 /note= "N-myristoylation site"  
 PT Domain 560..535 /note= "Transmembrane domain"  
 PT Region 627..633 /note= "N-myristoylation site"  
 PT Region 631..637 /note= "N-myristoylation site"  
 PT Region 640..646 /note= "N-myristoylation site"  
 PT Region 661..667 /note= "N-myristoylation site"  
 PT Region WO200218608-A2.  
 PN 07-MAR-2002.  
 PD 10-JUL-2001; 2001WO-US21635.  
 XX 29-AUG-2000; 2000US-228914P.  
 PR 11-JAN-2001; 2001US-0759056.  
 PR 10-JUL-2001; 2001US-0901812.  
 XX (GETH ) GENENTECH INC.  
 PA Tice D, Pennica D, Polakie P, Szeto W;  
 PI N-PSDB; ABA47278.  
 DR WPI: 2002-292202/33.  
 XX N-PSDB; ABA47278.  
 PT Selectively enhancing the expression of a protein in a tumour cell  
 PT characterised by aberrant Wnt signalling, useful for enhancing the  
 PT efficacy of treatment of cancers, comprising treating the tumour cell  
 PT with a retinoid -  
 PS Example 1; Fig 2; 141pp; English.  
 XX This invention relates to a novel method for selective enhancement of  
 CC the expression of a protein in a tumour cell characterised by aberrant  
 CC Wnt signalling by treating the tumour cell with a retinoid. The method  
 CC is useful for enhancing the efficacy of treatment of cancers  
 CC characterised by aberrant Wnt signalling, and for identifying tumour  
 CC antigens the expression of which is selectively enhanced by retinoid  
 CC treatment. The method is also useful in the early detection of  
 CC tumours for early intervention. The tumour antigens are useful as  
 CC targets for cancer therapy, since selective enhancement of their  
 CC expression by retinoid treatment relative to normal cells improves the  
 CC efficacy and therapeutic index of cancer therapeutics directed against  
 CC these antigens. Antibodies binding to the tumour antigens may be  
 CC administered for the treatment of various disorders or conditions,  
 CC including those characterised by overexpression and/or activation of the

CC tumour antigens, where such conditions or disorders include benign or  
 CC malignant tumours, leukaemias and lymphoid malignancies; neuronal, glial,  
 CC astrocytal, hypothalamic and other glandular, macrophagal, epithelial,  
 CC stromal and blastocytic disorders. The present sequence represents  
 CC the human strag protein (PRO10282) which is selectively upregulated by  
 CC retinoid using the method of the invention. The gene encoding the human  
 CC strag protein is located on human chromosome 15q23.  
 CC  
 XX Sequence 667 AA:  
 SQ  
 Query Match 98.8%; Score 1048; DB 23; Length 667;  
 Best Local Similarity 99.0%; Pred. No. 4.3e-106;  
 Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 2 SOPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLIL 61  
 Db 3 SQPAGNQTSPGATEDSYSGSWYIDEPQGXELQPEGEVPSCHTSIPGLYHACLASLIL 62  
 Oy 62 VLLLLMLVRRQLMPDCYRGPRGLPSVPDLADDRRAVPAVFWLLSLCLLPDD 121  
 Db 63 VLLLLMLVRRQLMPDCYRGPRGLPSVPDLADDRRAVPAVFWLLSLCLLPDD 122  
 Oy 122 ALPFLTLASAPSDGKTEAPRGAMKILGLFYALVLYPLAACATAGHTAHLIGSTLSWA 181  
 Db 123 ALPFLTLASAPSDGKTEAPRGAMKILGLFYALVLYPLAACATAGHTAHLIGSTLSWA 182  
 Oy 182 HLGQVQWORAECPOVPKI 199  
 Db 183 HLGQVQWORAECPOVPKI 200  
 RESULT 5  
 AAU83631  
 ID AAU83631 standard; Protein; 667 AA.  
 XX AC AAU83631;  
 XX DT 08-MAY-2002 (first entry)  
 XX DE Human PRO protein, Seq ID No 80.  
 XX KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;  
 KW breast cancer; prostate tumour; rectal tumour; liver tumour;  
 KW pituitary cell proliferation; chondrocyte cell proliferation;  
 KW tumour necrosis factor-alpha.  
 XX OS Homo sapiens.  
 XX PN WO200208288-A2.  
 XX PD 31-JAN-2002.  
 XX 29-JUN-2001; 2001WO-US21066.  
 XX 20-JUL-2000; 2000US-219556P.  
 XX 25-JUL-2000; 2000US-220585P.  
 XX 25-JUL-2000; 2000US-220605P.  
 XX 25-JUL-2000; 2000US-220607P.  
 XX 25-JUL-2000; 2000US-220624P.  
 XX 25-JUL-2000; 2000US-220638P.  
 XX 25-JUL-2000; 2000US-220664P.  
 XX 25-JUL-2000; 2000US-220666P.  
 XX 26-JUL-2000; 2000US-220893P.  
 XX 28-JUL-2000; 2000WO-US20710.  
 XX 23-AUG-2000; 2000WO-US23522.  
 XX 15-SEP-2000; 2000US-000000P.  
 XX 10-NOV-2000; 2000WO-US30873.  
 XX 28-NOV-2000; 2000US-253646P.  
 XX 01-DEC-2000; 2000WO-US32678.  
 XX 20-DEC-2000; 2000US-0747259.  
 XX 20-DEC-2000; 2000WO-US34956.  
 XX 26-FEB-2001; 2001WO-US06520.

PR 10-MAY-2001; 2001US-0854280.  
PR 25-MAY-2001; 2001WO-US17092.  
XX (GENTH ) GENENTECH INC.  
XX Baker KP, Deaneoyers L, Gerritsen ME, Goddard A, Godowski PJ,  
PI Grimaldi JC, Gurney AU, Smith V, Stephan JF, Watanabe CK, Wood WT,  
XX MPI; 2002-172001/22.  
DR N-PSDB; ABK33575.  
XX  
XX One hundred and twenty two nucleic acids encoding PRO polypeptides,  
PT useful for treating a PRO related disorder and for diagnosing tumours  
PT such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal  
PT tumour or liver tumour -  
XX  
XX Claim 11; Figure 80; 359pp; English.  
XX  
XX The invention relates to one hundred and twenty two nucleic acids  
CC encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides  
CC encode human secreted proteins. The PRO nucleic acids, polypeptides,  
CC agonists and antagonists are useful for treating a PRO related disorder.  
CC The PRO polypeptides are useful for diagnosing tumours, especially lung  
CC cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or  
CC liver tumour. The PRO polypeptides are useful for stimulating the  
CC proliferation of, or gene expression, in pericyte cells, for stimulating  
CC the proliferation or differentiation of chondrocyte cells, for  
CC stimulating the release of tumour necrosis factor-alpha from human blood,  
CC for stimulating or inhibiting the proliferation of normal human dermal  
CC fibroblast cells. The PRO polypeptide may also be used as molecular  
CC weight markers and for tissue typing. The PRO nucleic acids have  
CC applications in molecular biology, including use as hybridisation probes,  
CC and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO  
CC protein sequences of the invention.  
XX  
XX Sequence 667 AA:  
SQ  
Query Match 98.8%; Score 1048; DB 23; Length 667;  
Best Local Similarity 99.0%; Pred. No. 4.3e-106;  
Matches 196; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 2 SOPAGNOTSGATEDSYSGSWYIDPOGKXETOPGEPSCHTSIPIPGVTHACIASLIL 61  
DB 3 SOPAGNOTSGATEDSYSGSWYIDPOGSEBQPGEBVPSCHTSPRELTHACIASLIL 62  
QY 62 VLLIAMLVRRROJLMPDVCVRGPGCLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 121  
DB 63 VLLIAMLVRRROJLMPDVCVRGPGCLPSPVDFLAGDRPRAVPAVFMVLLSLCLLPDED 122  
QY 122 ALPPLTLASPSQDGKTEARPGAMKTLGLFHTAALTYELAACATAGHTAHLGSLTMA 181  
DB 123 ALPPLTLASPSQDGKTEARPGAMKTLGLFHTAALTYELAACATAGHTAHLGSLTMA 182  
QY 182 HLGVQVQWRAECPOVPKRI 199  
DB 183 HLGVQVQWRAECPOVPKRI 200  
RESULT 6  
ABB05048  
ID ABB05048 standard; Protein; 667 AA.  
XX  
AC ABB05048;  
XX  
DT 25-MAR-2002 (first entry)  
XX  
DB Human NOV8c protein SEQ ID NO:32.  
XX  
XX Human; NOVX; cytosratic; antidiabetic; anorectic; antibacterial;  
KW fungicide; virucide; protozoacide; analgesic; antiparkinsonian;  
KW antiaesthetic; hypotensive; osteoporotic; antiinflammatory; antifungal;  
KW neuroprotective; cardiant; antiallergic; antidepressant; nootropic;  
KW anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;

KW developmental disease; taste and scent detectability disorder; infection;  
KW Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;  
KW retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;  
KW noninsulin-dependent diabetes mellitus; acute heart failure; hypopension;  
KW hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;  
KW multiple sclerosis; angina pectoris; myocardial infarction; delirium;  
KW benign prostatic hypertrophy; manic depression; dementia; dyskinesia;  
KW severe mental retardation; Huntington's disease; chromosome 15;  
XX  
XX Gilles de la Tourette syndrome.  
OS Homo sapiens.  
XX  
XX WO200190155-A2.  
XX  
XX 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-US17073.  
XX  
XX 24-MAY-2000; 2000US-206679P.  
PR 24-MAY-2000; 2000US-206688P.  
PR 24-MAY-2000; 2000US-206828P.  
PR 30-MAY-2000; 2000US-207748P.  
PR 30-MAY-2000; 2000US-207798P.  
PR 31-MAY-2000; 2000US-208263P.  
PR 02-JUN-2000; 2000US-208831P.  
PR 05-JUN-2000; 2000US-209451P.  
PR 07-JUN-2000; 2000US-210060P.  
PR 20-JUL-2000; 2000US-219507P.  
PR 26-JUL-2000; 2000US-221337P.  
PR 31-JUL-2000; 2000US-221927P.  
PR 19-JAN-2001; 2001US-263135P.  
PR 24-JAN-2001; 2001US-263688P.  
PR 24-JAN-2001; 2001US-263694P.  
PR 23-MAY-2001; 2001US-0863776.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Spytek KA, Majumder K, Tchernev VT, Mishra V, Padigaru M,  
PI Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangolli E;  
XX WPI; 2002-106174/14.  
XX N-PSDB; ABA92669.  
DR  
XX Novel polypeptide, useful for treating pain, cancer, urinary retention,  
PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,  
PT asthma, ulcer, allergy and Huntington's disease, comprises isolated  
PT polypeptide NOVX -  
XX  
XX Claim 1; Page 106; 266pp; English.  
XX  
XX The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX  
CC proteins (I) have cytosratic, antidiabetic, anorectic, antibacterial,  
CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,  
CC antiaesthetic, hypotensive, osteoporotic, antiinflammatory, antitumor,  
CC neuroprotective, cardiant, antiallergic, antidepressant, nootropic, (II)  
CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)  
CC can be used in gene therapy and vaccine production. (I) and (II) can be  
CC used for treating or preventing a NOVX-associated disorder such as  
CC cardiomyopathy, atherosclerosis and diabetes in a human, where the  
CC disorder is related to cell signal processing and metabolic path way  
CC modulation, in a subject, preferably human. (I) and (II) can be used for  
CC diagnosing, preventing or treating developmental diseases, immune  
CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,  
CC signal transduction pathway disorders, retinal diseases including those  
CC involving photoreception, cell growth rate disorders, feeding disorders,  
CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,  
CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,  
CC Albrighr Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial  
CC infarction, allergies, benign prostatic hypertrophy, manic depression,  
CC delirium, dementia, severe mental retardation and dyskinesias, such as  
CC Huntington's disease or Gilles de la Tourette syndrome. The present  
CC sequence represents the human NOV8c protein from the present invention.

CC Human NOV2a is isolated to chromosome 15.  
 XX  
 SQ Sequence 667 AA;  
 Query Match 98.7%; Score 1047; DB 23; Length 667;  
 Best Local Similarity 98.5%; Pred. No. 5, 6e-106;  
 Matches 195; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 SOPAGNOTSGATEDSYSGSWYIDPQGGXELQEGEGVPSCHTSIPGGLYHACLASITL 61  
 DB 3 SOPAGNOTSGATEDSYSGSWYIDPQGGXELQEGEGVPSCHTSIPGGLYHACLASITL 62  
 QY 62 VLLILAMLYRRRLQMPDCCVGRPGCLPSVDFLAGDRPRAVPAVMVLLSSICLLLPDED 121  
 DB 63 VLLILAMLYRRRLQMPDCCVGRPGCLPSVDFLAGDRPRAVPAVMVLLSSICLLLPDED 122  
 QY 122 ALPFLTLASAPSGDKTEAPRGAMKILGLFYHVALYPLAACATAGTAHLLGSTLSMA 181  
 DB 123 ALPFLTLASAPSGDKTEAPRGAMKILGLFYHVALYPLAACATAGTAHLLGSTLSMA 182  
 QY 182 HLGQVQWQRAECPOVPKI 199  
 DB 183 HLGQVQWQRAECPOVPKI 200  
 RESULT 7  
 ABB05047  
 ID ABB05047 standard; Protein; 662 AA.  
 AC ABB05047;  
 XX 25-MAR-2002 (first entry)  
 DT  
 DE Human NOV8b protein SEQ ID NO:30.  
 XX  
 KW Human; NOVX; cytostatic; antidiabetic; anorectic; antibacterial;  
 KW fungicide; virucide; protozoacide; analgesic; antiparkinsonian;  
 KW antiaesthetic; hypotensive; osteopathic; antiinflammatory; antitumor;  
 KW neuroprotective; cardiant; antiallergic; antidepressant; nootropic;  
 KW anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;  
 KW developmental disease; taste and scent detectability disorder; infection;  
 KW Burkitt's lymphoma; signal transduction pathway disorder; cancer;  
 KW retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;  
 KW noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;  
 KW hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;  
 KW multiple sclerosis; angina pectoris; myocardial infarction; delirium;  
 KW benign prostatic hypertrophy; manic depression; dementia; dyskinesia;  
 KW severe metal retardation; Huntington's disease;  
 KW Gilles de la Tourette syndrome.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200190155-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 XX 24-MAY-2001; 2001WO-US17073.  
 PF  
 XX 24-MAY-2000; 2000US-206679P.  
 PR 24-MAY-2000; 2000US-206688P.  
 PR 24-MAY-2000; 2000US-206829P.  
 PR 30-MAY-2000; 2000US-207748P.  
 PR 30-MAY-2000; 2000US-207798P.  
 PR 31-MAY-2000; 2000US-208263P.  
 PR 02-JUN-2000; 2000US-208831P.  
 PR 05-JUN-2000; 2000US-209451P.  
 PR 07-JUN-2000; 2000US-21060P.  
 PR 20-JUL-2000; 2000US-219507P.  
 PR 26-JUL-2000; 2000US-221337P.  
 PR 31-JUL-2000; 2000US-221977P.  
 PR 19-JAN-2001; 2001US-263135P.  
 PR 24-JAN-2001; 2001US-263688P.  
 PR 24-JAN-2001; 2001US-263694P.

PR 23-MAY-2001; 2001US-0863776.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Spytek KA, Majumder K, Tchernev VT, Mishra V, Padigaru M;  
 PI Spederna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli E;  
 XX  
 XX WPI: 2002-106174/14.  
 DR N-FSDB; ABA92666.  
 DR  
 PT Novel polypeptide, useful for treating pain, cancer, urinary retention,  
 PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,  
 PT asthma, ulcer, allergy and Huntington's disease, comprises isolated  
 PT polypeptide NOVX -  
 XX  
 PS Claim 1; Page 103; 266pp: English.  
 XX  
 CC The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX  
 CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,  
 CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,  
 CC antiaesthetic, hypotensive, osteopathic, antiinflammatory, antitumor,  
 CC neuroprotective, cardiant, antiallergic, antidepressant, nootropic,  
 CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)  
 CC can be used in gene therapy and vaccine production (I) and (II) can be  
 CC used for treating or preventing a NOVX-associated disorder such as  
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the  
 CC disorder is related to cell signal processing and metabolic path way  
 CC modulation, in a subject, preferably human. (I) and (II) can be used for  
 CC diagnosing, preventing or treating developmental diseases, immune  
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,  
 CC signal transduction pathway disorders, retinal diseases including those  
 CC involving photoreception, cell growth rate disorders, feeding disorders,  
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,  
 CC Albright Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial  
 CC infarction, allergies, benign prostatic hypertrophy, manic depression,  
 CC delirium, dementia, severe metal retardation and dyskinesias, such as  
 CC Huntington's disease or Gilles de la Tourette syndrome. The present  
 CC sequence represents the human NOV8b protein from the present invention.  
 XX  
 SQ Sequence 662 AA;  
 Query Match 98.3%; Score 1043; DB 23; Length 662;  
 Best Local Similarity 98.5%; Pred. No. 1, 5e-105;  
 Matches 195; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 SOPAGNOTSGATEDSYSGSWYIDPQGGXELQEGEGVPSCHTSIPGGLYHACLASITL 61  
 DB 3 SOPAGNOTSGATEDSYSGSWYIDPQGGXELQEGEGVPSCHTSIPGGLYHACLASITL 62  
 QY 62 VLLILAMLYRRRLQMPDCCVGRPGCLPSVDFLAGDRPRAVPAVMVLLSSICLLLPDED 121  
 DB 63 VLLILAMLYRRRLQMPDCCVGRPGCLPSVDFLAGDRPRAVPAVMVLLSSICLLLPDED 122  
 QY 122 ALPFLTLASAPSGDKTEAPRGAMKILGLFYHVALYPLAACATAGTAHLLGSTLSMA 181  
 DB 123 ALPFLTLASAPSGDKTEAPRGAMKILGLFYHVALYPLAACATAGTAHLLGSTLSMA 182  
 QY 182 HLGQVQWQRAECPOVPKI 199  
 DB 183 HLGQVQWQRAECPOVPKI 200  
 RESULT 8  
 AAU91278  
 ID AAU91278 standard; Protein; 689 AA.  
 AC AAU91278;  
 XX  
 XX 18-JUN-2002 (first entry)  
 DT  
 XX Human NOV2b protein.



XX Human; NOXV; gene therapy; cardiomyopathy; atherosclerosis;  
 KW diabetes; cell signal processing; metabolic pathway modulation;  
 KW inflammation; autoimmune disorder; scleroderma; transplantation;  
 KW allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;  
 KW graft versus host disease; Leisch-Nyman syndrome; periodontitis;  
 KW pancreatitis; musculoskeletal disorder; Parkinson's disease;  
 KW Huntington's disease; behavioural disorder; pain; obesity; wound healing;  
 KW neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
 KW growth disorder; reproductive disorder; lung disease.  
 XX  
 OS Homo sapiens.  
 PN WO200216600-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 27-AUG-2001; 2001WO-US26518.  
 XX  
 PR 25-AUG-2000; 2000US-227800P.  
 PR 25-AUG-2000; 2000US-228205P.  
 PR 25-AUG-2000; 2000US-228324P.  
 PR 30-AUG-2000; 2000US-228397P.  
 PR 30-AUG-2000; 2000US-229185P.  
 PR 01-SEP-2000; 2000US-229780P.  
 PR 01-SEP-2000; 2000US-229848P.  
 PR 22-JAN-2001; 2001US-263337P.  
 PR 31-JAN-2001; 2001US-265318P.  
 PR 15-MAR-2001; 2001US-276451P.  
 PR 27-MAR-2001; 2001US-279196P.  
 PR 24-AUG-2001; 2001US-039339P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Gerlach V, McDougall JR, Smtison G, Stone DJ, Ellerman K;  
 PI Spytek KA, Zernusen BD, Rasceili L, Verney CAM, Patturajan M;  
 PI Tchernev VT, Padigan M, Taupier RJ;  
 XX  
 DR N-PSDB; ABR55562.  
 XX  
 PT MPI; 2002-292064/33.  
 XX  
 PT New isolated cytoplasmic, nuclear, membrane bound and secreted  
 PT polypeptides, termed NOXV, useful for treating inflammation, autoimmune  
 PT disorders, haemophilia, Leisch-Nyman syndrome, pancreatitis,  
 PT musculoskeletal disorders -  
 XX  
 PS Claim 1; Page 20; 245pp; English.  
 XX  
 CC The invention relates to an isolated cytoplasmic, nuclear, membrane bound  
 CC or secreted polypeptide, designated NOXV (actually NOXV, 2a, 2b 3a, 3b,  
 CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOXV, a  
 CC mature form, or a variant of the mature form of NOXV. Also included  
 CC are a polynucleotide encoding NOXV (or its complement), a vector  
 CC comprising the polynucleotide, a cell comprising the vector, an  
 CC anti-NOXV antibody, determining the presence of NOXV in a sample  
 CC using the antibody, determining the presence of NOXV polynucleotide in  
 CC a sample using a probe which binds to NOXV polynucleotide, identifying a  
 CC agent which binds to NOXV (including modulators of NOXV). NOXV, the  
 CC polynucleotide and the antibody are useful for diagnosing, treating  
 CC or preventing a NOXV-associated disorder selected from cardiomyopathy,  
 CC atherosclerosis, diabetes, a disorder related to cell signal processing,  
 CC and metabolic pathway modulation, inflammation, autoimmune disorders,  
 CC scleroderma, transplantation, allergies, systemic lupus erythematosus,  
 CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,  
 CC Leisch-Nyman syndrome, periodontitis, pancreatitis, musculoskeletal  
 CC disorders, Parkinson's disease, Huntington's disease, behavioural  
 CC disorders, pain, neurodegenerative and neuropsychiatric disorders,  
 CC hypertension, wound healing, obesity, growth and reproductive  
 CC disorders, lung diseases and many other diseases and disorders listed in  
 CC the specification. NOXV, the polynucleotide and the antibody are useful  
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue  
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assay,

CC prognostic assays, monitoring clinical trials and pharmacogenomic), and  
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOXV is  
 CC useful as immunogen to produce antibodies immunospecific for NOXV, as  
 CC vaccines to screen for potential agonist and antagonist compounds, and  
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide  
 CC is useful in gene therapy, to express NOXV, to detect NOXV RNA  
 CC or a genetic lesion in a NOXV gene, and to modulate NOXV activity. The  
 CC vector is useful for producing non-human transgenic animals. The antibody  
 CC is useful for isolating, and purifying NOXV and to monitor protein levels  
 CC in tissue as part of a clinical testing procedure. The present sequence  
 CC represents a NOXV protein.  
 XX  
 SO Sequence 689 AA;  
 Query Match 96.1%; Score 1020; DB 23; Length 689;  
 Best Local Similarity 88.6%; Pred. No. 5.3e-103;  
 Matches 195; Conservative 1; Mismatches 2; Indels 22; Gaps 1;  
 QY 2 SOPAGNQTSPGATEDYSGSWYIDPEQGXELQPEGEVPSCHSIPPGYLHACLAASLIL 61  
 DB 3 SOPAGNQTSPGATEDYSGSWYIDPEQGXELQPEGEVPSCHSIPPGYLHACLAASLIL 62  
 QY 62 VLLLLMLVRRRLQMPDCVGRPGLP-----SPVDFLAGDRPR 99  
 DB 63 VLLLLMLVRRRLQMPDCVGRPGLPSEWVGQGLNSLILPCPCLCASFVDFLAGDGR 122  
 QY 100 AVPAAVFMTLSICLLLPDEBALPCLTLASPSODGKTPAPGAWKITGLFYALALYP 159  
 DB 123 AVPAAVFMTLSICLLLPDEBALPCLTLASPSODGKTPAPGAWKITGLFYALALYP 182  
 QY 160 LAACATAGHTAAHLGSLTSMWHLGQVWQRAECPOVPKTI 199  
 DB 183 LAACATAGHTAAHLGSLTSMWHLGQVWQRAECPOVPKTI 222  
 XX  
 RESULT 9  
 AAU04558  
 ID AAU04558 standard; Protein: 658 AA.  
 XX  
 AC AAU04558;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE Human Str66 homologue, PRO19578.  
 XX  
 KW Human; Str66, PRO19578; 15q23; tumour; breast cancer;  
 KW head tumour; leukaemia; lymphoid malignancy; inflammatory disorder;  
 KW immunogenic disorder; antigen; antibody.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 8..12  
 FT /note= "Asn is N-glycosylated"  
 FT Region 50..56  
 FT /label= N\_myristoylation\_site  
 FT Domain 54..71  
 FT /label= Transmembrane\_domain  
 FT Domain 93..111  
 FT /label= Transmembrane\_domain  
 FT Region 123..131  
 FT /note= "ATP/GTP-binding site motif A (P-loop)"  
 FT Domain 140..157  
 FT /label= Transmembrane\_domain  
 FT Region 167..173  
 FT /label= N\_myristoylation\_site  
 FT Domain 197..214  
 FT /label= Transmembrane\_domain  
 FT Region 232..238  
 FT /label= N\_myristoylation\_site  
 FT Domain 291..312  
 FT /label= Transmembrane\_domain  
 FT Region 308..314

FT /label= N\_myristoylation\_site  
 FT 332..338  
 FT /label= N\_myristoylation\_site  
 FT 355..366  
 FT /note= "Prokaryotic membrane lipoprotein attachment site"  
 FT 356..371  
 FT /label= Transmembrane\_domain  
 FT 425..444  
 FT /label= Transmembrane\_domain  
 FT 464..481  
 FT /label= Transmembrane\_domain  
 FT 505..522  
 FT /label= Transmembrane\_domain  
 FT 516..522  
 FT /label= N\_myristoylation\_site  
 FT 618..624  
 FT /label= N\_myristoylation\_site  
 FT 622..628  
 FT /label= N\_myristoylation\_site  
 FT 631..637  
 FT /label= N\_myristoylation\_site  
 FT 652..658  
 FT /label= N\_myristoylation\_site

XX MO200151635-A2.  
 XX 19-JUL-2001.  
 XX 11-JAN-2001, 2001WO-US00847.  
 XX 13-JAN-2000, 2000US-0175849.  
 XX 14-APR-2000, 2000US-0197089.  
 XX 29-AUG-2000, 2000US-0228914.  
 XX (GETH ) GENENTECH INC.  
 XX Pennica D, Smith V, Wood WI;  
 XX N-PSDB; AAS08632.  
 DR WPI; 2001-442146/47.  
 XX Human nucleic acid encoding a PRO19578 polypeptide (the PRO polypeptides are referred to as Str6 polypeptides), useful in molecular biology, including use as hybridisation probes, and in chromosome and gene mapping -  
 XX Claim 24, Fig 7; 15pp; English.  
 CC The sequence is human PRO19578 (a variant possibly due to alternative splicing of the Str6 mRNA), a homologue of murine Str6, a retinoic acid responsive protein thought to play an important role in early dorsoventral limb patterning during development and later in the control of endochondral ossification. The gene for the Str6 homologue is located on chromosome 15q23. The Str6 polypeptides, agonists, antagonists or anti-str6 antibodies are useful for preparation of a medicament useful in the treatment of a condition which is responsive to the Str6 polypeptides, agonists, antagonists or anti-Str6 antibodies. The Str6 polypeptides may also be employed as molecular weight markers for protein electrophoresis. The Str6 nucleic acids have applications in molecular biology, including use as hybridisation probes, and in chromosome and gene mapping. The antibodies and other anti-tumour compounds may be used to treat various conditions, including those characterised by overexpression and/or activation of amplified genes. Exemplary conditions or disorders include benign or malignant tumours (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulval, thyroid, hepatic carcinomas, sarcomas, glioblastomas, and various head and neck tumours, leukemias and lymphoid malignancies), neuronal, glial, astrocytal, haemochromatosis and other glandular, macrophagal, epithelial, stromal and blastococytic disorders, inflammatory, angiogenic and immunogenic disorders.

SO Sequence 658 AA:  
 Query Match 93.0%; Score 986.5; DB 22; Length 658;  
 Best Local Similarity 94.4%; Pred. No. 2,4e-99;  
 Matches 187; Conservative 1; Mismatches 9; Gaps 1;  
 QY 2 SOPAGNOTSGATEDSYSGSWYIDEPQGXELQPEGVPSCHTSIPRGLTHACLASLIL 61  
 DB 3 SOPAGNOTSGATEDSYSGSWYIDEPQGBELQPEGVPSCHTSIPGLTHACLASLIL 62  
 QY 62 VLLLLMLVRRRLQMPDCVGRGPGPLSPVDFLAGDPRPAVAFAVWLLSLCLLPDED 121  
 DB 63 VLLLLMLVRRRLQMPDCVGRGPGPLSPVDFLAGDPRPAVAFAVWLLSLCLLPDED 113  
 QY 122 ALPPIITLASAPSGDKTEAPRGAMKILGLPHYALVYPLAACATAGHTAAHLGSLTSM 181  
 DB 114 ALPPIITLASAPSGDKTEAPRGAMKILGLPHYALVYPLAACATAGHTAAHLGSLTSM 173  
 QY 182 HLGQVQWQRAECQVPEKI 199  
 DB 174 HLGQVQWQRAECQVPEKI 191

## RESULT 10

AAU78576  
 ID AAU78576 standard; Protein; 658 AA.

AC AAU78576;

DT 18-JUN-2002 (first entry)

DE Human Str6 protein variant (PRO19578).

XX Human cancer; Str6; PRO19578; cytostatic; stromal disorder;  
 KW tumour; retinoid; Wnt; tumour antigen; Leukaemia; lymphoid malignancy;  
 KW neuronal disorder; glial disorder; astrocytal disorder;  
 KW hypochalamic disorder; glandular disorder; macrophagal disorder;  
 KW epithelial disorder; blastococic disorder; chromosome 15q23.  
 XX Homo sapiens.

OS Homo sapiens.  
 FH Key  
 FT 8..12 Location/Qualifiers  
 FT /note= "N Glycosylation site"  
 FT 50..56  
 FT /note= "N-myristoylation site"  
 FT 54..71  
 FT /note= "Transmembrane domain"  
 FT 93..111  
 FT /note= "Transmembrane domain"  
 FT 123..131  
 FT /note= "ATP/GTP binding site motif A"  
 FT 140..157  
 FT /note= "Transmembrane domain"  
 FT 167..173  
 FT /note= "N-myristoylation site"  
 FT 197..214  
 FT /note= "Transmembrane domain"  
 FT 212..238  
 FT /note= "N-myristoylation site"  
 FT 291..312  
 FT /note= "Transmembrane domain"  
 FT 308..314  
 FT /note= "N-myristoylation site"  
 FT 332..338  
 FT /note= "N-myristoylation site"  
 FT 356..371  
 FT /note= "Transmembrane domain"  
 FT 355..366  
 FT /note= "lipoprotein lipid attachment site"  
 FT 425..444  
 FT /note= "Transmembrane domain"  
 FT 464..481

```

FT      Region /note= "Transmembrane domain"
FT      516..522
FT      /note= "N-myristoylation site"
FT      Domain 505..522
FT      /note= "Transmembrane domain"
FT      Region 618..628
FT      /note= "N-myristoylation site"
FT      Region 622..628
FT      /note= "N-myristoylation site"
FT      Region 631..637
FT      /note= "N-myristoylation site"
FT      Region 652..658
FT      /note= "N-myristoylation site"
FT      Region 652..658
FT      /note= "N-myristoylation site"
XX      WO200218608-A2.
XX      07-MAR-2002.
XX      10-JUL-2001; 2001WO-US21635.
XX      29-AUG-2000; 2000US-228914P.
XX      11-JAN-2001; 2001US-0759056.
XX      10-JUL-2001; 2001US-0901812.
XX      (GENTH ) GENENTECH INC.
XX      Tice D, Pennica D, Polakis P, Szeto W;
XX      WPI; 2002-292202/33.
XX      N-PSDB; ABR47280.
XX      Selectively enhancing the expression of a protein in a tumour cell
XX      characterised by aberrant Wnt signalling, useful for enhancing the
XX      efficacy of treatment of cancers, comprises treating the tumour cell
XX      with a retinoid
XX      Example 1; Fig 7; 14pp; English.
XX      This invention relates to a novel method for selective enhancement of
XX      the expression of a protein in a tumour cell characterised by aberrant
XX      Wnt signalling by treating the tumour cell with a retinoid. The method
XX      is useful for enhancing the efficacy of treatment of cancers
XX      characterised by aberrant Wnt signalling, and for identifying tumour
XX      antigens the expression of which is selectively enhanced by retinoid
XX      treatment. The method is also useful in the early detection of
XX      tumours for early intervention. The tumour antigens are useful as
XX      targets for cancer therapy, since selective enhancement of their
XX      expression by retinoid treatment relative to normal cells improves the
XX      efficacy and therapeutic index of cancer therapeutics directed against
XX      these antigens. Antibodies binding to the tumour antigens may be
XX      administered for the treatment of various disorders or conditions,
XX      including those characterised by overexpression and/or activation of the
XX      tumour antigens, where such conditions or disorders include benign or
XX      malignant tumours; leukaemias and lymphoid malignancies; neuronal, glial,
XX      astrocytal, hypochalamic and other glandular, macrophagal, epithelial,
XX      stromal and blascoeloid disorders. The present sequence represents
XX      the human Str6 protein variant (PRO19578) which is actively upregulated
XX      by retinoid using the method of the invention. The gene encoding the
XX      human Str6 protein is located on human chromosome 15q23.
XX      Sequence 658 AA:
SQ
Query Match 93.0%; Score 986.5; DB 23; Length 658;
Best Local Similarity 94.4%; Pred. No. 2.4e-99;
Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;
QY 2 SOPAGNQTSPGATSDYSGVSWYIDBPQGXLOPEGVPSCHTSIPRGLVHACIASLSTL 61
DB 3 SOPAGNQTSPGATSDYSGVSWYIDBPQGXLOPEGVPSCHTSIPRGLVHACIASLSTL 62
QY 62 VLLLLAMLVRRQLWPDVCVRGRLPSPVDLADGRPAVPAVFWLLSSLCLLLPDD 121
DB 63 VLLLLAMLVRRQLWPDVCVRGRLPSPVDLADGRPAVPAVFWLLSSLCLLLPDD 113

```

```

QY 122 ALPELTLASAPSDGKTEAPRGAMKILGLFYHALLVYPLAACATAGTAAHLGSLTSLWA 181
DB 114 ALPELTLASAPSDGKTEAPRGAMKILGLFYHALLVYPLAACATAGTAAHLGSLTSLWA 173
QY 182 HLGQVWORAECGPVPKI 199
DB 174 HLGQVWORAECGPVPKI 191
RESULT 11
ABB05046
ID ABB05046 standard; Protein; 653 AA.
XX
AC ABB05046;
XX
DT 25-MAR-2002 (first entry)
XX
DE Human NOV8a protein SEQ ID NO:28.
XX
KW Human; NOVX; cytosolic; antidiabetic; anorectic; antibacterial;
KW fungicide; virucide; protozoacide; analgesic; antiparkinsonian;
KW antistatic; hypotensive; osteopathic; antiinflammatory; antitumor;
KW neuroprotective; cardiac; antiallergic; antidepressant; nootropic;
KW anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;
KW developmental disease; taste and scent detectability disorder; infection;
KW Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;
KW retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;
KW noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;
KW hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;
KW multiple sclerosis; angina pectoris; myocardial infarction; delirium;
KW benign prostatic hypertrophy; manic depression; dementia; dykinesia;
KW severe mental retardation; Huntington's disease; chromosome 15;
KW Gilles de la Tourette syndrome.
XX
OS Homo sapiens.
XX
PN WO200190155-A2.
XX
PD 29-NOV-2001.
XX
PF 24-MAY-2001; 2001WO-US17073.
XX
PR 24-MAY-2000; 2000US-206679P.
XX
PR 24-MAY-2000; 2000US-206688P.
XX
PR 24-MAY-2000; 2000US-206829P.
XX
PR 30-MAY-2000; 2000US-207748P.
XX
PR 30-MAY-2000; 2000US-207798P.
XX
PR 31-MAY-2000; 2000US-208263P.
XX
PR 02-JUN-2000; 2000US-208831P.
XX
PR 05-JUN-2000; 2000US-209451P.
XX
PR 07-JUN-2000; 2000US-210060P.
XX
PR 20-JUL-2000; 2000US-219507P.
XX
PR 26-JUL-2000; 2000US-221337P.
XX
PR 31-JUL-2000; 2000US-221927P.
XX
PR 19-JAN-2001; 2001US-26135P.
XX
PR 24-JAN-2001; 2001US-263688P.
XX
PR 24-JAN-2001; 2001US-263694P.
XX
PR 23-MAY-2001; 2001US-0863776.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Spytek KA, Majumder K, Tchernev VT, Mishra V, Padigaru M;
PI Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangoli E;
XX
DR WPI; 2002-106174/14.
XX
N-PSDB; ABR2663.
XX
PT Novel polypeptide, useful for treating pain, cancer, urinary retention,
PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
PT actna, ulcer, allergy and Huntington's disease, comprises isolated
PT polypeptide NOVX -
XX

```

PS Claim 1, Page 98; 266pp; English.

XX The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX  
CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,  
CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,  
CC antiepileptic, hypotensive, osteoporosis, antiinflammatory, antitumor,  
CC neuroprotective, cardiatic, antiallergic, antidepressant, nootropic,  
CC anticonvulsant and neuroleptic activities (I) and polynucleotides (II)  
CC can be used in gene therapy and vaccine production. (I) and (II) can be  
CC used for treating or preventing a NOVX-associated disorder such as  
CC cardiovascular, atherosclerosis and diabetes in a human, where the  
CC disorder is related to cell signal processing and metabolic path way  
CC modulation, in a subject, preferably human. (I) and (II) can be used for  
CC diagnosis, preventing or treating developmental diseases, immune  
CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,  
CC signal transduction pathway disorders, retinal diseases including those  
CC involving photoreception, cell growth rate disorders, feeding disorders,  
CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,  
CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,  
CC Albright Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial  
CC infarction, allergies, benign prostatic hypertrophy, manic depression,  
CC delirium, dementia, severe mental retardation and dyskinesias, such as  
CC Huntington's disease or Gilles de la Tourette syndrome. The present  
CC sequence represents the human NOV9a protein from the present invention.  
XX Human NOV9a is isolated to chromosome 15.

SO Sequence 653 AA:

Query Match 92.9%; Score 985.5; DB 23; Length 653;  
Best Local Similarity 94.4%; Pred. No. 3,18-99;  
Matches 187; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

QY 2 SQPAGNCTSGATGEDSYGSGWYIDEPQGXELQPEGEFSPCHTSPGLYHACLASLTL 61

DB 3 SQPAGNCTSGATGEDSYGSGWYIDEPQGXELQPEGEFSPCHTSPGLYHACLASLTL 62

QY 62 VLLLLMLVRRQLPMDPCVGRGRPGULSPVDFLAGDRPAVPAVFWLLSLCLLLPDED 121

DB 63 VLLLLMLVRRQLPMDPCVGRGRPGULSPVDFLAGDRPAVPAVFWLLSLCLLLPDED 122

QY 122 ALPPLTASAPSDQGTETAPRAWKTLGLFHYAALYPLACATRGTAHLLGSLTSM 181

DB 123 ALPPLTASAPSDQGTETAPRAWKTLGLFHYAALYPLACATRGTAHLLGSLTSM 173

QY 182 HLGVOVQWRAECPOVPK 199

DB 174 HLGVOVQWRAECPOVPK 191

RESULT 12

AAU91277

ID AAU91277 standard; Protein; 660 AA.

XX AAU91277;

DT 18-JUN-2002 (first entry)

XX Human NOV9a protein.

XX Human; NOVX; gene therapy; cardiomyopathy; atherosclerosis;

XX diabetes; cell signal processing; metabolic pathway modulation;

XX inflammation; autoimmune disorder; scleroderma; transplantation;

XX allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;

XX graft versus host disease; Leach-Nyhan syndrome; peridontitis;

XX pancreatitis; musculoskeletal disorder; Parkinson's disease;

XX Huntington's disease; behavioural disorder; pain; obesity; wound healing;

XX neurodegenerative disorder; neuropsychiatric disorder; hypertension;

XX growth disorder; reproductive disorder; lung disease.

XX Homo sapiens.

XX WO200216600-A2.

XX 28-FEB-2002.

XX 27-AUG-2001; 2001WO-US26518.

XX 25-AUG-2000; 2000US-227800P.

XX 25-AUG-2000; 2000US-228205P.

XX 25-AUG-2000; 2000US-228324P.

XX 30-AUG-2000; 2000US-228997P.

XX 30-AUG-2000; 2000US-229185P.

XX 01-SEP-2000; 2000US-229780P.

XX 01-SEP-2000; 2000US-229848P.

XX 01-SEP-2000; 2000US-229850P.

XX 22-JAN-2001; 2001US-263137P.

XX 31-JAN-2001; 2001US-265518P.

XX 15-MAR-2001; 2001US-276451P.

XX 27-MAR-2001; 2001US-279196P.

XX 24-AUG-2001; 2001US-039339P.

XX (CURA-) CURAGEN CORP.

XX Gerlach V, MacDougall JR, Smithson G, Stone DJ, Ellerman K;

XX Sytek KA, Zehmsen BD, Rastelli L, Verney CM, Paturajan M;

XX Tchernov VT, Padigar M, Taupier RJ;

XX WPI: 2002-292064/33.

XX N-PDB; ABK55561.

XX New isolated cytoplasmic, nuclear, membrane bound and secreted

XX polypeptides, termed NOVX, useful for treating inflammation, autoimmune

XX disorders, haemophilia, Leach-Nyhan syndrome, pancreatitis,

XX musculoskeletal disorders -

XX Claim 1; Page 18; 245pp; English.

XX The invention relates to an isolated cytoplasmic, nuclear, membrane bound

XX or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,

XX 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a

XX mature form, or a variant of the mature form of NOVX. Also included

XX are a polynucleotide encoding NOVX (or its complement), a vector

XX comprising the polynucleotide, a cell comprising the vector, an

XX anti-NOVX antibody, determining the presence of NOVX in a sample

XX using the antibody, determining the presence of NOVX polynucleotide in

XX a sample using a probe which binds to NOVX polynucleotide, identifying a

XX an agent which binds to NOVX (including modulators of NOVX). NOVX, the

XX polynucleotide and the antibody are useful for diagnosing, treating

XX or preventing a NOVX-associated disorder selected from cardiomyopathy,

XX atherosclerosis, diabetes, a disorder related to cell signal processing

XX and metabolic pathway modulation, inflammation, autoimmune disorders,

XX scleroderma, transplantation, allergies, systemic lupus erythematosus,

XX haemophilia, graft versus host disease, Alzheimer's disease, stroke,

XX Leach-Nyhan syndrome, peridontitis, pancreatitis, musculoskeletal

XX disorders, Parkinson's disease, Huntington's disease, behavioural

XX disorders, pain, neurodegenerative and neuropsychiatric disorders,

XX hypertension, wound healing, obesity, growth and reproductive

XX disorders, lung diseases and many other diseases and disorders listed in

XX the specification. NOVX, the polynucleotide and the antibody are useful

XX in screening assays, detection assays (e.g., chromosomal mapping, tissue

XX typing, forensic biology), predictive assays (e.g., diagnostic assays,

XX prognostic assays, monitoring clinical trials and pharmacogenomic), and

XX in methods of treatment (e.g., therapeutic and prophylactic). NOVX is

XX useful as immunogen to produce antibodies immunospecific for NOVX, as

XX vaccines to screen for potential agonist and antagonist compounds, and

XX as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide

XX is useful in gene therapy, to express NOVX, to detect NOVX mRNA

XX or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The

XX vector is useful for producing non-human transgenic animals. The antibody

XX is useful for isolating, and purifying NOVX and to monitor protein levels

XX in tissue as part of a clinical testing procedure. The present sequence

XX represents a NOVX protein.

XX Sequence 660 AA:

Query Match 84.6%; Score 898; DB 23; Length 660;  
 Best Local Similarity 89.0%; Pred. No. 1,2e-89;  
 Matches 178; Conservative 1; Mismatches 5; Indels 16; Gaps 4;

QY 2 SGPAGNQTSGATEDSYSGSWITDEPOGXELOPGEVPSCHTSI PGLYHACIASISFQ 59  
 DB 3 SGPAGNQTSGATEDSYSGSWITDEPOGXELOPGEVPSCHTSI PGLYHACIASISFQ 61

QY 60 ILVLLTAMLVRRRQIMPCVGRCPGLPSPVDFLAGDRPRAVPAVFMVLLSSICLLLP 119  
 DB 62 ILVLLTAMLVRRRQIMPCVGRCPGLPSPVDFLAGDRPRAVPAVFMVLLSSICLLLP 117

QY 120 EDALPFTTLTASASODGKTEAPRGAMKILGFHYALYYPLAACATGHTAAHLISSTLS 179  
 DB 118 EDALPFTTLTASASODGKTEAPRGAMKILGFHYALYYPLAACATGHTAAHLISSTLS 168

QY 180 MAHLGVQVWQRAECPOVPKI 199  
 DB 169 MAHLGVQVWQRAECPOVPKI 188

RESULT 13  
 AAB94108  
 ID AAB94108 standard; Protein; 560 AA.  
 XX  
 AC AAB94108;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:14340.  
 XX  
 KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 XX  
 PR 27-AUG-1999; 99JP-0300253.  
 XX  
 PR 11-JAN-2000; 2000JP-0118776.  
 XX  
 PR 02-MAY-2000; 2000JP-0183767.  
 XX  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR MPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 full-length cDNAs defined in the specification, and for the detection  
 and/or diagnosis of the abnormality of the proteins encoded by the  
 full-length cDNAs -  
 PT  
 PS Claim 8; SEQ ID 14340; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 full-length cDNAs defined in the specification. Where a primer set  
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 to the complementary strand of a polynucleotide which comprises one of  
 the 5602 nucleotide sequences defined in the specification, where the  
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 of an oligonucleotide comprising a sequence complementary to the  
 complementary strand of a polynucleotide which comprises a 5'-end  
 sequence and an oligonucleotide comprising a sequence complementary to a  
 polynucleotide which comprises a 3'-end sequence, where the  
 polynucleotide comprises at least 15 nucleotides and the combination of  
 the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and  
 in gene therapy. The primers are useful for synthesizing polynucleotides,  
 particularly full-length cDNAs. The primers are also useful for the  
 detection and/or diagnosis of the abnormality of the proteins encoded by  
 the full-length cDNAs. The primers allow obtaining of the full-length  
 cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 of the present invention.

XX  
 SQ Sequence 560 AA;  
 Query Match 46.0%; Score 488; DB 22; Length 560;  
 Best Local Similarity 98.9%; Pred. No. 1e-44; Indels 0; Gaps 0;  
 Matches 92; Conservative 1; Mismatches 0;

QY 107 MVLSSICLLPDEDALPFTLTASASODGKTEAPRGAMKILGFHYALYYPLAACATA 166  
 DB 1 MVLSSICLLPDEDALPFTLTASASODGKTEAPRGAMKILGFHYALYYPLAACATA 60

QY 167 GHTAAHLISSTLSMAHLGVQVWQRAECPOVPKI 199  
 DB 61 GHTAAHLISSTLSMAHLGVQVWQRAECPOVPKI 93

RESULT 14  
 AABG02155  
 ID AABG02155 standard; Protein; 1070 AA.  
 XX  
 AC AABG02155;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #2146.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR MPI; 2001-639362/73.  
 XX  
 DR N-PSDB; AAB66342.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PT  
 PS Claim 20; SEQ ID No 32514; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences: (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (ii) and its binding partners are useful in medical  
CC imaging of sites expressing (ii). (i) and (ii) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostic, forensic, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG0377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 1070 AA;

Query Match 8.2%; Score 87.5; DB 22; Length 1070;  
Best Local Similarity 25.1%; Pred. No. 2.1;  
Matches 50; Conservative 12; Mismatches 70; Indels 67; Gaps 7;

QY 7 NOTSPGATEDYSGSWYID---EPQGXELQEGEVPSCHTSIIPGLYH--ACTASLSIL 61  
DB 633 NLDPNASEDYTSSEYQFNGELPGRHKIPESQFPFSTLSISGSLFPVACGS----- 687  
QY 62 VLLLLALVRRQLMPGRCGLPSPVDFLAGDRPA----- 100  
DB 688 VLSEGSPLTEQESSFSDHRSKTVSAS---TGDLPRASKIDEPALOTANASAPLVPA 743  
QY 101 ---VPAVFMVLISICLLLPEDALPFLTLASAPS-----ODGK 137  
DB 744 VMNCTPCPSMVLICLSMLTRSVYSPHSPHSSPSSTAVSVTVESVHPTCTSTIOEAK 803  
QY 138 TEA-----PRGAWKI 147  
DB 804 TRAADLVNPLDPNADKI 822

RESULT 15

AA48063 ID AAR48063 standard; Protein; 380 AA.

AA48063;

20-JUL-1994 (first entry)

Sequence of polypeptide encoded by the first open reading frame in  
DE the unique short (Us) region of bovine herpes virus (BHV) genome.

XX Insertion region; unique short region; Us; vaccine; antigen.

XX Bovine herpes virus type 1, strain ST.

XX MO9400586-A.

XX 06-JAN-1994.

XX 25-JUN-1993; 93MO-FR00642.

XX 26-JUN-1992; 92FR-0007930.

XX (IMWR ) RHONE MERIEUX SA.

XX Audomnet JF, Legasletois ICMA, Leung-lack P, Riviere MEA;

XX MPI; 1994-026222/03.

XX N-PSDB; AA053350.

PT New insertion region sequence of bovine herpes virus genomic DNA  
PT - used for recombinant virus with this region deleted or  
PT inactivated, useful in vaccines allowing differentiation between  
PT vaccinated and infected cattle

XX Claim 9; Fig 2; 47pp; French.

CC The 4190 bp sequence in AA053350 encodes polypeptides homologous to  
CC HSV-1 gE, gB and gS; these are neither essential for in vitro  
CC replication nor important in inducing a protective immune response.  
CC It is genomic DNA purified from the ST strain of BHV-1. (BHV 1B also  
CC known as infectious bovine rhinotracheitis virus.) The sequence  
CC forms the insertion region in genomic DNA. xBHV in which the  
CC specific insertion region, esp. nucleotides 172-1311, has been  
CC deleted or inactivated by insertion are claimed.

SO Sequence 380 AA;

Query Match 8.1%; Score 85.5; DB 15; Length 380;  
Best Local Similarity 27.6%; Pred. No. 0.95;  
Matches 37; Conservative 12; Mismatches 38; Indels 47; Gaps 6;

QY 4 PAGNOTSPGATEDYSGSWYIDEPQGXELQEGEVPSCHTSIIPGLYHACLAS---LS 59  
DB 256 PAQPTASPAASPSRAFA---AAPAAAPADDTA-----RFRQLASILVPLC 302  
QY 60 ILVLLIAM-----LVRRRL-----WPDYGRGRLPSPV 90  
DB 303 VLVLALLALCAATVNCALRRRLPCSRVRYKRTCAACGSGTCAGRPCKGAAPSAATV 362  
QY 91 DFLAGDRPRAVPA 104  
DB 363 VAL-GSRPRAPPLA 375

Search completed: March 13, 2003, 11:28:52  
Job time: 29.0649 secs

*Handwritten notes:*  
"The 4190 bp sequence in AA053350 encodes polypeptides homologous to HSV-1 gE, gB and gS; these are neither essential for in vitro replication nor important in inducing a protective immune response. It is genomic DNA purified from the ST strain of BHV-1. (BHV 1B also known as infectious bovine rhinotracheitis virus.) The sequence forms the insertion region in genomic DNA. xBHV in which the specific insertion region, esp. nucleotides 172-1311, has been deleted or inactivated by insertion are claimed."  
"us-09-816-653a-2.rag"  
"Page 12"

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:26:32 ; Search time 34.1126 Seconds  
(without alignments)  
1963.067 Million cell updates/sec

Title: US-09-816-653A-4  
Perfect score: 1695  
Sequence: 1 CYSALVLSCLTFLVLMRS.....APPTLWESOGCFWRKLVG 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mmc:\*
- 8: sp\_organellae:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriopl:\*
- 17: sp\_archaeopl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1110.5	65.5	667	4	Q9BX79 homo sapien
2	1110.5	65.5	667	4	Q9BX79 homo sapien
3	1106.5	65.3	658	4	Q9BX78 homo sapien
4	832.5	49.1	670	11	Q70491 mus musculu
5	133	7.8	621	11	Q9DBN1 mus musculu
6	98.5	5.8	1841	11	Q9ER60 mus musculu
7	97.5	5.8	403	4	Q75205 homo sapien
8	96.5	5.7	593	16	Q83179 treponema p
9	92	5.4	796	3	Q9P3B4 neosporea
10	92	5.4	2008	4	Q96S58 homo sapien
11	92	5.4	2146	4	Q9XR73 homo sapien
12	92	5.4	2146	4	Q9XR73 homo sapien
13	91.5	5.4	581	1	Q07289 nactonomora
14	91	5.4	379	8	Q8WDS8 lepidosiren
15	90.5	5.3	229	8	Q37632 rhagoletis
16	90.5	5.3	229	8	Q37636 rhagoletis

17	90.5	5.3	320	8	Q8WA20 ophiophagus
18	90.5	5.3	400	2	Q8VNT4 enterobacte
19	90.5	5.3	730	12	Q06634 bovine herp
20	90	5.3	418	4	Q8RDP0 homo sapien
21	90	5.3	414	4	Q9BUE5 homo sapien
22	90	5.3	476	4	Q9BUE5 homo sapien
23	89.5	5.3	107	16	Q8X7Y2 escherichia
24	89.5	5.3	318	8	Q8WA25 matricora bl
25	89.5	5.3	319	8	Q8WA18 pseudasplis
26	89.5	5.3	361	16	Q8XTH6 raletonia s
27	89	5.3	1082	10	Q8WTH0 sorghum bic
28	89	5.3	3154	12	Q69088 human herpe
29	88.5	5.2	374	8	Q21210 parachromis
30	88.5	5.2	395	16	Q8Z119 salmonella
31	88.5	5.2	402	16	Q8ZMH0 salmonella
32	87.5	5.2	226	8	Q20419 hodoctermops
33	87.5	5.2	229	8	Q9ZXZ2 panulirus a
34	87.5	5.2	229	8	Q37663 rhagoletis
35	87.5	5.2	229	8	Q37439 carpomya sh
36	87.5	5.2	318	8	Q8WA21 maja nivea
37	87.5	5.2	367	4	Q96M97 homo sapien
38	87.5	5.2	377	8	Q21104 archocentru
39	87.5	5.2	463	10	Q9ASAO oryza sativ
40	87.5	5.2	554	6	Q95L89 bos taurus
41	87.5	5.2	580	6	Q8WML4 bos taurus
42	87	5.1	2015	4	Q96J69 homo sapien
43	87	5.1	2159	11	Q91V24 mus musculu
44	86.5	5.1	294	2	Q05135 pseudomnas
45	86.5	5.1	301	16	Q915M4 pseudomnas

## ALIGNMENTS

### RESULT 1

Q9BX79 PRELIMINARY; PRT; 667 AA.

AC Q9BX79; Q9BX98; 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 21, Last annotation update)

DE STRA6 isoform 1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Szeto W., Jiang W., Tice D.A., Rubinfield B., Hollingshead P.G.,

RA Fong S.E., Dugger D.L., Pham T., Yansura D.G., Wong T.A.,

RA Grimaldi J.C., Corpuz R.T., Singh J.S., Frantz G.D., Devaux B.,

RA Crowley C.W., Schwall R.H., Eberhard D.A., Rastelli L., Polakis P.,

RA Pennica D.D.

RT Overexpression of the retinoic acid-responsive gene Stra6 in human

RT cancers and its synergistic induction by Mnt-1 and retinoic acid."

RL Cancer Res. 0:0-0(2001).

RN [2]

RP SEQUENCE OF 108-667 FROM N.A.

RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Nagatsuna M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takahashi M., Chiba Y., Ichida S., Murakawa K., Ono Y., Takiguchi S.,

RA Watanabe S., Kimura K., Morakami K., Ichii S., Kawai Y., Saito K.,

RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,

RA Ninomiya K., Iwayanagi T.,

RT "NEBO human cDNA sequencing project."

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF527278; AAK30289.1. -

DR EMBL; AK022603; BAB14122.1. -

FT VARIANT 527 M -> I.

SO SEQUENCE 667 AA; 73502 MW; D20840A6998BA2E CRC64;

Query Match 65.5%; Score 1110.5; DB 4; Length 667;

Best Local Similarity 76.6%; Pred. No. 4.9e-90;  
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSGSGD-----GQFSWNLFS 42  
Db 365 CYISALVLSCLTFLVLMRSVLTNRALHRAALDLSPLHRSPPHROAIFCWSFS 424

QY 43 VPLRLPPLAGLVQOIIFFLGTALAFVLMRVLHGRLFFRSLESSPFWLTALAVI 102  
Db 425 AVGTATFCGLLVQOIIFFLGTALAFVLMRVLHGRLFFRSLESSPFWLTALAVI 484

QY 103 LONMAAHWFLETHDGHQPLTNRRVLYAATFLLPPLVLYGAMXXXC----- 149  
Db 485 LONMAAHWFLETHDGHQPLTNRRVLYAATFLLPPLVLYGAMVATWVLLSALYNAIHL 544

QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATFCSLLAQOSLLPRTMAAPD 203  
Db 545 GQMDLSLPRRAATLDPGYTYRNFLKIEVSQSHPMATFCSLLAQOSLLPRTMAAPD 604

QY 204 SLRPGDEEGMQLLOTKDSMAKARPGARARWGLAYTLHNPTLQVFRKTALLGANG 263  
Db 605 SLRPGDEEGMQLLOTKDSMAKARPGARARWGLAYTLHNPTLQVFRKTALLGANG 664

QY 264 AOP 266  
Db 665 AOP 667

RESULT 2  
Q8TB21  
ID Q8TB21 PRELIMINARY; PRT; 667 AA.  
AC Q8TB21;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical 73.5 kDa protein.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxId=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Streusberg R.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC025256; AAK25256.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 667 AA; 73532 MW; 120685F4635CB0E5 CRC64;

Query Match 65.5%; Score 110.5; DB 4; Length 667;  
Best Local Similarity 76.6%; Pred. No. 4.9e-90;  
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSGSGD-----GQFSWNLFS 42  
Db 365 CYISALVLSCLTFLVLMRSVLTNRALHRAALDLSPLHRSPPHROAIFCWSFS 424

QY 43 VPLRLPPLAGLVQOIIFFLGTALAFVLMRVLHGRLFFRSLESSPFWLTALAVI 102  
Db 425 AVGTATFCGLLVQOIIFFLGTALAFVLMRVLHGRLFFRSLESSPFWLTALAVI 484

QY 103 LONMAAHWFLETHDGHQPLTNRRVLYAATFLLPPLVLYGAMXXXC----- 149  
Db 485 LONMAAHWFLETHDGHQPLTNRRVLYAATFLLPPLVLYGAMVATWVLLSALYNAIHL 544

QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATFCSLLAQOSLLPRTMAAPD 203  
Db 545 GQMDLSLPRRAATLDPGYTYRNFLKIEVSQSHPMATFCSLLAQOSLLPRTMAAPD 604

QY 204 SLRPGDEEGMQLLOTKDSMAKARPGARARWGLAYTLHNPTLQVFRKTALLGANG 263  
Db 605 SLRPGDEEGMQLLOTKDSMAKARPGARARWGLAYTLHNPTLQVFRKTALLGANG 664

QY 264 AOP 266  
Db 665 AOP 667

RESULT 3  
Q9BX78  
ID Q9BX78 PRELIMINARY; PRT; 658 AA.  
AC Q9BX78;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE STRA6 isoform 2.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxId=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Szeo W., Jiang W., Tice D.A., Rubinfeld B., Hollingshead P.G.,  
RA Fong S.E., Dugger D.L., Pham T., Yansura D.G., Wong T.A.,  
RA Grimaldi J.C., Corpuz R.T., Singh J.S., Frantz G.D., Devaux B.,  
RA Crowley C.W., Schwall R.H., Eberhard D.A., Rastelli L., Polakis P.,  
RA Benicla D.;  
RT Overexpression of the retinoic acid-responsive gene Stra6 in human  
RT cancers and its synergistic induction by wnt-1 and retinoic acid.";  
RL Cancer Res. 0:0-0(2001).  
DR EMBL; AF352729; AAK30290.1; -.  
FT VARIANT 518 518 I -> M.  
SQ SEQUENCE 658 AA; 72582 MW; 5405281727F647DD CRC64;

Query Match 65.3%; Score 1106.5; DB 4; Length 658;  
Best Local Similarity 76.2%; Pred. No. 1.1e-89;  
Matches 231; Conservative 5; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSGSGD-----GQFSWNLFS 42  
Db 356 CYISALVLSCLTFLVLMRSVLTNRALHRAALDLSPLHRSPPHROAIFCWSFS 415

QY 43 VPLRLPPLAGLVQOIIFFLGTALAFVLMRVLHGRLFFRSLESSPFWLTALAVI 102  
Db 416 AVGTATFCGLLVQOIIFFLGTALAFVLMRVLHGRLFFRSLESSPFWLTALAVI 475

QY 103 LONMAAHWFLETHDGHQPLTNRRVLYAATFLLPPLVLYGAMXXXC----- 149  
Db 476 LONMAAHWFLETHDGHQPLTNRRVLYAATFLLPPLVLYGAVATWVLLSALYNAIHL 535

QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATFCSLLAQOSLLPRTMAAPD 203  
Db 536 GQMDLSLPRRAATLDPGYTYRNFLKIEVSQSHPMATFCSLLAQOSLLPRTMAAPD 595

QY 204 SLRPGDEEGMQLLOTKDSMAKARPGARARWGLAYTLHNPTLQVFRKTALLGANG 263  
Db 596 SLRPGDEEGMQLLOTKDSMAKARPGARARWGLAYTLHNPTLQVFRKTALLGANG 655

QY 264 AOP 266  
Db 656 AOP 658

RESULT 4  
O70491  
ID O70491 PRELIMINARY; PRT; 670 AA.  
AC O70491;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE Retinoic acid-responsive protein.  
GN STRA6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;



RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9537753; PubMed=7649373;  
RA Bouillet P., Oulied-Abdelghani M., Vicaire S., Garnier J.M.,  
RA Schaubert B., Dolle P., Chambon P.  
RT "Efficient cloning of cDNAs of retinoic acid-responsive genes in p19  
RT embryonal carcinoma cells and characterization of a novel mouse gene,  
RT Stral (mouse LERK-2/Epi92)."  
RL Dev. Biol. 170:420-433(1995).  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9746723; PubMed=9203140;  
RA Bouillet P., Sapin V., Chazaud C., Mesasded N., Decimo D., Dolle P.,  
RA Chambon P.;  
RT "Developmental expression pattern of Stra6, a retinoic acid-responsive  
RT gene encoding a new type of membrane protein."  
RL Mech. Dev. 63:173-186(1997).  
RN  
RP SEQUENCE FROM N.A.  
RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF062475; AAC16016.1; -  
DR MED: MG1:107742; Stra6  
DR MED: MG1:107742; Stra6  
SQ SEQUENCE 670 AA; 73774 MW; 6AC6E9F3BE16A4F7 CRC64;  
  
Query Match 49.1%; Score 832.5; DB 11; Length 670;  
Best Local Similarity 57.7%; Pred. No. 2e-65;  
Matches 176; Conservative 28; Mismatches 62; Indels 39; Gaps 4;  
  
QY 1 CYISAVLSGLTFLVMSLVYHR-----LGSQSGSD-----GQFSWNLFS 42  
DB 366 CYISAVLSGLTFLVMSLVYHR-----LGSQSGSD-----GQFSWNLFS 425  
QY 43 VPLPLPLAGLIVQIIFFLGTTALAFVLMPLVHGRNLFPSLESSENPFLTALAVI 102  
DB 426 AVOTAFSCGLLAVQVIFFLGTTSLAFVFPVPLHGRNLLRLESTWFFMLTALAVI 485  
QY 103 LONAAHWFLETHDHPOLTNRRVLYATFLPLVPLVGMKXKCSPIAIRHPT-- 159  
DB 486 LONAAHWFLETHDHPOLTNRRVLYATFLPLVPLVGMKXKCSPIAIRHPT-- 159  
QY 160 -----PGYTYRNFLKIEVSQSHPMATFCSLLQAQSLPRMAPOD 203  
DB 546 GOMDLILPORAASLDPGYHYTQNFRIASOSHPGVAFPCALLHAPSPPRPPLAPD 605  
QY 204 SLRPEDEGQGLIQTQDSNAKAPGAPXGRARGLAVTLLHNPVLQVFRKTLALG-A 261  
DB 606 SLRPEDEGQGLIQTQDSNAKAPGAPXGRARGLAVTLLHNPVLQVFRKTLALG 665  
QY 262 NGAQP 266  
DB 666 NGTOP 670  
  
RESULT 5  
Q9DBN1 PRELIMINARY; PRT; 621 AA.  
ID Q9DBN1  
AC Q9DBN1  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)  
DE 1300002K09Rik protein.  
CN 1300002K09Rik  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxId=10090;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=LIVER;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nihi K., Kiyosawa H., Kondo S., Yamana K. I.,  
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Resole G., Quackenbush J.,  
RA Schirral L.M., Stabl P., Suzuki R., Tomita M., Wagner L., Wainio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guenichon S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mazzarelli J., Monhaerts P.,  
RA Norton P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Saeeki H., Sato K., Schoenbach C., Seta Y., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
DR EMBL: AK004855; BAB23619.1; -  
DR MED: MG1:1921402; 1300002K09Rik  
SQ SEQUENCE 621 AA; 71002 MW; 909E1C6BC2A9EA60 CRC64;  
  
Query Match 7.8%; Score 133; DB 11; Length 621;  
Best Local Similarity 23.1%; Pred. No. 0.0015;  
Matches 55; Conservative 37; Mismatches 92; Indels 54; Gaps 7;  
  
QY 50 LAGLIVQIIFFLGTTALAFVLMPLVHGRNLFPSLESSENPFLTALAV---ILQNM 106  
DB 388 LAGLIVQIIFFLGTTALAFVLMPLVHGRNLFPSLESSENPFLTALAV---ILQNM 444  
QY 107 AAHWFLETHDHPD-----LITNRVLYATFLPLVPLVGMKXKCSPIAIR 155  
DB 445 IAGTFPLQKLTQKSDKQPLANNRRAHNFNFLEFFYVVLG--LGAISLRLISCLLG 502  
QY 156 -----RHPTPGYTYRNFLKIEVSQSHPMATFCSLLQAQSLPRMA 199  
DB 503 TWLIARIDPTTQSGSYBADMGPFAMQMLFVHYHTNPVJVSFCIIILTS----- 553  
QY 200 APDLSRPEDEGQGLIQTQDSNAKAPG-AXGRARGLAVTLLHNPVLQVFRKT 256  
DB 554 -----HKDRKLQKTVKXKCLNOSAGPRFSARATRFVLTLLINPRVLMLKS 602  
  
RESULT 6  
Q9ER60 PRELIMINARY; PRT; 1841 AA.  
ID Q9ER60  
AC Q9ER60  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)  
DE Voltage-gated sodium channel.  
CN SCN4A.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxId=10090;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=HEART;  
RA Zimmer T., Beandorf K.;  
RT "The mouse heart sodium channels: cloning and characterization of two  
RT distinct isoforms and alternatively spliced variants."  
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ278787; CAC17146.1; -  
DR MED: MG1:98250; Scn4a.  
DR InterPro: IPR001682; Ca/Na\_pore.  
DR InterPro: IPR000048; IC\_region.  
DR InterPro: IPR000636; M+channel\_nlg.  
DR InterPro: IPR001496; Na\_channel.  
DR InterPro: IPR003915; PKD\_2.  
DR Pfam: PF00520; Ion\_transf\_4.  
DR Pfam: PF00612; IQ\_1.

DR PRINTS: PRO0170; NACHANNEL.  
 DR PRINTS: PRO1433; POLYCISTIN2.  
 DR SMART: SM00015; IQ; 1.  
 DR Ionic channel.  
 SQ SEQUENCE 1841 AA; 208797 MW; 076DFD33A9E0E55 CRC64;

Query Match 5.8%; Score 98.5; DB 11; Length 1841;  
 Best Local Similarity 22.0%; Pred. No. 5.4;  
 Matches 67; Conservative 35; Mismatches 105; Indels 97; Gaps 14;

QY 35 QPSNLEVPPLPLAGLVQIIFLGTALAFVLPVHGNLLFFSLSSSPFW 94  
 DB 633 QGNMIFP-----SFIYTLISVELG---LANYGSLSVSRFLRLVFLKSNP-T 679  
 QY 95 LTLALAVILQNMMAHWFLETHDGHPTLRRLVLAATFLPL--NYLVGAMXXXCSPS 152  
 DB 680 LNMILIKIIGNSVGA-----LSNLTIVLAILVIFRAVVGMLPKRSYKECVCK 726  
 QY 153 IAINRPTGYTT---YRNPLKI-----EVSQHPAMTAFCSLLLOAQ-- 191  
 DB 727 IASDGLPRHMHDFHSLIVFRILCGEMITMDQMEVAGQAMCLTVFLMVAIGNLV 786  
 QY 192 -----SLPRTMAAPDLSRPGEDGMOQLQTKDSMAKGPAGKGRARWGLAVT-- 243  
 DB 787 VLNLFLALLISFSA--DSLASDEDEGMNNIQ-----IANGIKMGIAPAKT 832  
 QY 244 ----LHNPTLOVRKTKALL-----GANGAQCSSLPGSPSITPMAQPPPHXG 291  
 DB 833 FLGLGLGKLSL--KQIMLSLQPGAGENGESP-----PEDEKKEPPEDEG 878  
 QY 292 XVEY 295  
 DB 879 NKEI 882

## RESULT 7

075205 PRELIMINARY; PRT; 403 AA.

AC 075205; 09NZHO; 08, Created)  
 DT 01-NOV-1998 (Tremblrel). 08, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel). 21, Last annotation update)  
 DE Hypochemical 44.8 kDa protein (GPCRS protein) (Orphan G-protein coupled receptor).  
 GN A-96912.1 OR GPCRSB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Loftus B.J., Zhou L., Crosby M., Fuhmann J., Mason T.M.,  
 RA Brandon R., Kim U.J., Kerlavage A.R., Venter J.C.,  
 RT "Homo sapiens Chromosome 16 BAC clone C17875C-A-6912.";  
 RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Wei H., Osborne B., Spruyt M., Murphy D.;  
 RT "Cloning of a novel G protein-coupled receptor localized on human  
 RT chromosome 16p12.";  
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Robbins M.J., Michalovich D., Hill J., Calver A.R., Medhurst A.D.,  
 RA Cloger I., Sime M.A., Middlemiss D.N., Pangalos M.N.;  
 RT "Molecular cloning and characterisation of two novel retinoic acid  
 RT inducible orphan G-protein coupled receptors.";  
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 9-403 FROM N.A.  
 RX MEDLINE=20247251; PubMed=10783259;  
 RA Brauner-Osborne H., Krosgaard-Larsen P.;

RT "Sequence and expression pattern of a novel human orphan G-protein-  
 RT coupled receptor, GPCRSB, a family C receptor with a short amino-  
 RT terminal domain.";  
 RL Genomics 65:121-128(2000).  
 DR EMBL: AC004131; AAC7544.1;  
 DR EMBL: AF181862; AAF05331.1;  
 DR EMBL: AF202640; CAC00632.1;  
 DR EMBL: AF202640; AAF67321.1;  
 DR Interpro: IPR002956; Bride of 7less.  
 DR Interpro: IPR000337; GPCR\_Mgr.  
 DR Pfam: PF00003; 7tm\_3; 1.  
 DR PRINTS: PRO1223; BRIDDE07LESS.  
 DR PROSITE: PS50259; G\_PROTEIN\_RECEP\_F3\_4; 1.  
 KW Hypochemical protein; Receptor.  
 SQ SEQUENCE 403 AA; 44795 MW; 3902A16C4F69C26E CRC64;

Query Match 5.8%; Score 97.5; DB 4; Length 403;  
 Best Local Similarity 19.9%; Pred. No. 1.3;  
 Matches 69; Conservative 30; Mismatches 107; Indels 14; Gaps 14;

QY 5 ALVSLCLTFELVLMASLVTHRLGSGSGDGFSLNLSVPLPLIAGLLVQII----- 59  
 DB 134 ALCPSCLLSQARVRLVRH-----GTCPAGQVGLALCL---MLVGVIIAWEVL 181  
 QY 60 -----FLGTALAFVLMVMP-----VLHGNLLFFRS 86  
 DB 182 VLTVRDTPACAYEPMDFVMAIIVDWLVLTGLALFTLCQPKRWKLNCAFLITAF 241  
 QY 87 LE-SWPFMLTL-----ALAVILQNMMAHWFLETHDGHPTLRRLV 128  
 DB 242 LSVLIWVAMMTWYLFQNVKLGQGDAMDPTLAILTL--ASGVVFLFH-ALPEI----- 292  
 QY 129 YAATFLPLFNLVVGAMKXXCSPSIAIRHPTPGYTYRNFLEKIEVSGHPANTAFCSLL 188  
 DB 293 -----HCTLLPALQENTPNVF-----DTSQPMRETAE----- 320  
 QY 189 QASQLPRTMAPODSLPGEDGMOQLQTKDSMAKGPAGKGRARWGLAVTLNMP 248  
 DB 321 EEDVQLPRAYV-----ENKAFMDENNAALTAAPPGSLCK----- 357  
 QY 249 TLQVRKTKLLGANGAQCSSLPGSPSITPMAQ-----PAGPPHXG 291  
 DB 358 -----RPSGLKGRSAPFRSNVYQPTEMAVVINGTITPAPPSHTG 399

## RESULT 8

083179 PRELIMINARY; PRT; 593 AA.

AC 083179;  
 DT 01-NOV-1998 (Tremblrel). 08, Created)  
 DT 01-NOV-1998 (Tremblrel). 08, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel). 19, Last annotation update)  
 DE Thiamine ABC transporter, permease protein, putative.  
 GN TP0143.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 NCBI\_Taxid=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NICHOLS;  
 RX MEDLINE=9633770; PubMed=9665876;  
 RA Frazer C.M., Norris S.J., Weinatlock G.M., White O., Sutton G.G.,  
 RA Dodeon R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,  
 RA McDonald L., Arifach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 DR EMBL: AE001199; AAC65131.1; -;  
 DR TIGR; TP0143; -;

DR InterPro; IPR000515; BPD\_transp.  
 DR Pfam; PF00528; BPD\_transp; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 593 AA; 64882 MW; 5D2E3F9189A9FCB CRC64;

Query Match 5.7%; Score 96.5; DB 16; Length 593;  
 Best Local Similarity 23.0%; Pred. No. 2.4;  
 Matches 61; Conservative 29; Mismatches 86; Indels 89; Gaps 12;

QY 3 ISALVLSCLT-----TFVLMKSLVTHRLGSGSGDQGFSNVLPFPLIAGLIVQOI 58  
 DB 120 LGSALVACVGGAAFFVKKSLRALALFPA-----LCALPLALPPL----- 161  
 QY 59 IFFLGTTALAPVLMFVHGNLLFFSLSSWPF-----MTLLALA 100  
 DB 162 -----TMAALFLFST-----GKGGCTRLGLMGVSTPRTFLXSAGVIAHAMNPLA 212  
 QY 101 VILQMAAMWVLETHDHPQLTNR--RVLYATFELLFPLNVLVGMXXX-----CSP 151  
 DB 213 LATIARAAMTISADAEADARLGAAVAVFTTIT-----LPALAGLRSSFLVIFLYCFP 267  
 QY 152 SIAT-----RHPTPGYTYRNFLKIEVSGSHP-----AMTAFCSLLT----- 188  
 DB 268 SLMMVLLIGCTFTTLEVELYRS--ITQAAHPYASTALSGTFALLMINGVSTBESQ 324  
 QY 189 QAQSLPRTMAAPODSLRPEDEG 213  
 DB 325 QAQSCVARTRPLPTQSIQNTIRIG 349

## RESULT 9

Q9P3B4 PRELIMINARY; PRT; 796 AA.

AC Q9P3B4; 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Related to ser/thr protein kinase IME2.  
 GN B7N4.100.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schulte U., Aign V., Hohnesel J., Brandt P., Fattmann B., Holland R.,  
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RU [2]  
 RP SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RA Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AL390218; CAB99241.2; -  
 DR HSSP; P24941; 1BUH.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; UNKNOWN\_1.  
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE\_SF; 2.  
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 KM SEQUENCE 796 AA; 87388 MW; 6981C17DBEBB97AB CRC64;

Query Match 5.4%; Score 92; DB 3; Length 796;

Best Local Similarity 20.7%; Pred. No. 8.3; Indels 118; Gaps 14;  
 Matches 75; Conservative 38; Mismatches 131;  
 QY 23 THRLLSGGS-----GDGQFQSWLPSVPLPLPLAGLIVQO--IFFLGTTALAF 69

DB 13 SHGIGSQQLAEDEFVLEKIEIGSGSRG-----SVVLAFVRSAGATVARGVTIAIKTKKTF 68

QY 70 LVMPVLHGNLLFFSLSS-----WPFLLTALAVLQMAAMWVLETHG 118

DB 69 ESVGPCMELEREVVFLKTLPAHPLPALDIFLDP--TKLHIAEMYMESGNLYOLMKARD 126

QY 119 HPQLTNRV-----LYATFV--LPLNVLVGMXXXCSPSIAIRH----- 157

DB 127 HKCIDNSSVSIILFQIMKGHEIHAAHFFHRD IKENILVSTSSHMATNSFRYAGALNN 186

QY 158 --PTPGYTYRNFLKIEVSGSHPANTAPFCSLLQASLLP-----RTMAAPODSLRPG 208

DB 187 PPTPTPTTYTK-----IADFLARETHSKLPTTYVSTRKRADEVILRAG 232

QY 209 EEDGGMOL-----LOTDSMAKGRPGAXGRARWGLAVTLNHPLOVFRKTAALGA-- 261

DB 223 EYSAFVDIMAIQMAVAEIALIKPLPGG-----NEVDQVRVEIIMSGSPG 277

QY 262 -----NGAOPCSLFGSPSITP-----AMQAPSPDHGXVEVCLHW 299

DB 278 NWYNKAGARVGGEGREGRTRLAGKLGSPFKKAPHSMDTIILQTFQWASLAHFVTCWKW 337

QY 300 EP 301

DB 338 DP 339

## RESULT 10

Q96S58 PRELIMINARY; PRT; 2008 AA.

AC Q96S58; 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE ABCA-SSN.  
 GN ABCA7/ABCA-SSN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21255283; PubMed=11355874;  
 RA Tanaka A., Ikeda Y., Abe-Dohmae S., Arakawa R., Sadanami K.,  
 RA Kiketa A., Nakagawa S., Nagase T., Aoki R., Kiketa N., Anachi T.,  
 RA Yokoyama S., Ueda K.;  
 RT "Human ABCA1 Contains a Large Amino-Terminal Extracellular Domain  
 RT Homologous to an Epitope of Sjogren's Syndrome."  
 RL Biochem. Biophys. Res. Commun. 283:1019-1025(2001).  
 DR EMBL; AB053390; BAB62294.1; -  
 DR InterPro; IPR004339; ABC\_transport.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR Pfam; PF00005; ABC\_tran; 2.  
 DR ProDom; PD000006; ABC\_transporter; 2.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 SQ SEQUENCE 2008 AA; 218617 MW; 226FF85C2420B90 CRC64;

Query Match 5.4%; Score 92; DB 4; Length 2008;  
 Best Local Similarity 25.3%; Pred. No. 22;  
 Matches 80; Conservative 48; Mismatches 126; Indels 62; Gaps 17;

QY 3 ISALVLSCLT-----LTV-----LVLMKSLVTHRLGSGSGD-----QFQSWNL--FSV 43

DB 1396 VDIVAVSICVAFMSFVPSFTLVLEERVTRAKHLQIMGLSTFLYMGFLMDNCNTLV 1455

QY 44 PLPLPLAGLIVQOIIFFLGTTALAPVLMFVHGNLL-----FFSLSSWPFWLT 96

DB 1456 PACIVVLIIFLAGQORA--YVAPANLPLALLLILLYGWSITPLMYPASFFFSVSTAYVVLT 1514

QY 97 -LALAVLQMAAMWVLETHDHPQLTNRVLAATFELLFPLNVL--VGMAMXXXCSPSIA 154

Db 1515 CINLFTGINGSMAATFVLELPSDOKLOEVS--ILKOVFLIPPHCLGGLIDMVRNOAMA 1572  
 Qy 155 IRHPTGYTYTYRNFLKIEVSOSH-----PAMTAFCSLLLOAOSLL--PRTMAAPOD 203  
 Db 1573 DAFERLGDROFOSPLRMEVVGKILAMVIGQPLFLF--TLLQHSQQLPQPRVSLP-- 1629  
 Qy 204 SLRPGEDGEGQQLLOTDSMAKARGP-----AXRGARGLAYTLLHNPTLOVF 253  
 Db 1630 --LIGEBDE--DVARERERVQATGCDVLYLNLTKYVRGQRMAYVRLCLGIPGECF 1685  
 Qy 254 RKTALLGANGAQPCSS 269  
 Db 1686 ---GLGVNGAGKST 1698

RESULT 11  
 Q9NR73 PRELIMINARY; PRT; 2146 AA.  
 AC Q9NR73;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Macrophage ABC transporter.  
 GN ABCA7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2034305; PubMed=10873640;  
 RA Kaminski W.E., Orso E., Diederich W., Klucken J., Drobnik W.,  
 RA Schmitz G.,  
 RT "Identification of a Novel Human Sterol-Sensitive ATP-Binding Cassette  
 RT Transporter (ABCA7)."  
 RL Biochem. Biophys. Res. Commun. 273:532-538(2000).  
 DR EMBL: AF250238; AAF85794.1; -  
 DR InterPro: IPR003593; AAA ATPase.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR Pfam: PF00005; ABC\_tran; 2.  
 DR ProDom: PD000006; ABC\_transport; 2.  
 DR SMART: SM00382; AAA; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 KW ATP-binding.  
 SQ SEQUENCE 2146 AA; 234469 MW; 679816B2D75FFD CRC64;

Query Match 5.4%; Score 92; DB 4; Length 2146;  
 Best Local Similarity 25.3%; Pred. No. 24;  
 Matches 80; Conservative 48; Mismatches 126; Indels 62; Gaps 17;

Qy 3 ISALVLSCL---LTF-----LVLMRSLV---HRLGSGSGD-----GQFSNVL--FSV 43  
 Db 1534 VDVLVSLICVVFAMSFVPASFTVLIERVYTRAKQLQLMGSLPTLYWLGFLMDMCNYLV 1593  
 Qy 44 PLPLPLAGLLVQOIIFFLGTALAFVLPVPLHGRNLL-----FPRSLESSMPFWLT 96  
 Db 1594 PACTIVLILFAPQRA-YVAPANLPALLLLLLLGKSTPLMYRPAFFSFVSTAYVLT 1652  
 Qy 97 -LALAVILQNMMAHWVLETHDHPQLTNRVLYAATFLPLVNL-VGAMXXXXCSPSIA 154  
 Db 1653 CINLFTGINGSMAATFVLELPSDOKLOEVS--ILKOVFLIPPHCLGGLIDMVRNOAMA 1710  
 Qy 155 IRHPTGYTYTYRNFLKIEVSOSH-----PAMTAFCSLLLOAOSLL--PRTMAAPOD 203  
 Db 1711 DAFERLGDROFOSPLRMEVVGKILAMVIGQPLFLF--TLLQHSQQLPQPRVSLP-- 1767  
 Qy 204 SLRPGEDGEGQQLLOTDSMAKARGP-----AXRGARGLAYTLLHNPTLOVF 253  
 Db 1768 --LIGEBDE--DVARERERVQATGCDVLYLNLTKYVRGQRMAYVRLCLGIPGECF 1823  
 Qy 254 RKTALLGANGAQPCSS 269

Db 1824 ---GLGVNGAGKST 1836

RESULT 12  
 Q9BZC4 PRELIMINARY; PRT; 2146 AA.  
 AC Q9BZC4;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE ABC transporter member 7.  
 GN ABCA7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21328888; PubMed=11435699;  
 RA Brocardo C., Osorio J., Luciano M.-F., Schrieml L.M., Pradee C.,  
 RA Shulman S., Arnold I., Naudin L., Lafargue C., Rosier M., Jordan B.,  
 RA Matei M.G., Dean M., Deneffe P., Chimi G.,  
 RT "Comparative analysis of the promoter structure and genomic  
 RT organization of the human and mouse ABCA7 gene encoding a novel ABCA  
 RT transporter."  
 RL Cytogenet. Cell Genet. 92:264-270(2001).  
 DR EMBL: AF328787; AAK0959.1; -  
 DR InterPro: IPR003593; AAA ATPase.  
 DR InterPro: IPR003439; ABC\_Transport.  
 DR Pfam: PF00005; ABC\_tran; 2.  
 DR ProDom: PD000006; ABC\_transport; 2.  
 DR SMART: SM00382; AAA; 2.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
 DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 KW ATP-binding.  
 SQ SEQUENCE 2146 AA; 234306 MW; 2391728D5AD97875 CRC64;

Query Match 5.4%; Score 92; DB 4; Length 2146;  
 Best Local Similarity 25.3%; Pred. No. 24;  
 Matches 80; Conservative 48; Mismatches 126; Indels 62; Gaps 17;

Qy 3 ISALVLSCL---LTF-----LVLMRSLV---HRLGSGSGD-----GQFSNVL--FSV 43  
 Db 1534 VDVLVSLICVVFAMSFVPASFTVLIERVYTRAKQLQLMGSLPTLYWLGFLMDMCNYLV 1593  
 Qy 44 PLPLPLAGLLVQOIIFFLGTALAFVLPVPLHGRNLL-----FPRSLESSMPFWLT 96  
 Db 1594 PACTIVLILFAPQRA-YVAPANLPALLLLLLLGKSTPLMYRPAFFSFVSTAYVLT 1652  
 Qy 97 -LALAVILQNMMAHWVLETHDHPQLTNRVLYAATFLPLVNL-VGAMXXXXCSPSIA 154  
 Db 1653 CINLFTGINGSMAATFVLELPSDOKLOEVS--ILKOVFLIPPHCLGGLIDMVRNOAMA 1710  
 Qy 155 IRHPTGYTYTYRNFLKIEVSOSH-----PAMTAFCSLLLOAOSLL--PRTMAAPOD 203  
 Db 1711 DAFERLGDROFOSPLRMEVVGKILAMVIGQPLFLF--TLLQHSQQLPQPRVSLP-- 1767  
 Qy 204 SLRPGEDGEGQQLLOTDSMAKARGP-----AXRGARGLAYTLLHNPTLOVF 253  
 Db 1768 --LIGEBDE--DVARERERVQATGCDVLYLNLTKYVRGQRMAYVRLCLGIPGECF 1823  
 Qy 254 RKTALLGANGAQPCSS 269  
 Db 1824 ---GLGVNGAGKST 1836

RESULT 13  
 Q07289 PRELIMINARY; PRT; 581 AA.  
 AC Q07289;  
 DT 01-JUN-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Subunit I-homologue of cytochrome oxidase.  
 GN CBAA.  
 OS Naeromonas pharaonis (Naeromonas bacterium pharaonis).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Naeromonas.  
 OX NCBI\_TaxID=2257;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SP1/28;  
 RX MEDLINE=98088558; PubMed=9428682;  
 RA Mactar S., Engelhard M.;  
 RT "Cytochrome b3 from Naeromonas bacterium pharaonis: An archaeal four-subunit cytochrome-c-type oxidase."  
 RL Eur. J. Biochem. 250:332-341(1997).  
 CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).  
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
 DR EMBL: Y10500; CAU71525.1; -;  
 DR HSSP: Q56408; IEHK.  
 DR InterPro: IPR000883; COX1.  
 DR Pfam: PF00115; COX1; 1.  
 KW Copper; Heme; Oxidoreductase; Respiratory chain; Transmembrane.  
 SQ SEQUENCE 581 AA; 63990 MW; 46963269A92D0522 CRC64;

Query Match 5.4%; Score 91.5; DB 1; Length 581;  
 Best Local Similarity 26.6%; Pred. No. 6.6;  
 Matches 63; Conservative 32; Mismatches 89; Indels 53; Gaps 15;

QY 6 LVLSCLTFLVYMRSLVTRHLSGGSGGDFSNVLPPLPLPLAGLVQIIIFLGT 65  
 DB 300 LILPLSLTFTVVAAMEHGARGGSG--GYFGM-LRALPMRDVPTGMALAGLMF-----A 352  
 QY 66 ALAFVLMVHGRNLLPFRSLSSW---PFMLTALAVLONMAAHVFLETHGHQ 121  
 DB 353 AAFFSGM-VNAGMINTL-VHNTWVVGPHLTVGTAVALTFMAVSWEFL-----PQ 402  
 QY 122 LTNRR-----VLVAATFLFL-PLNVLGAMKXCCSPSIAIRHPGYVTRNFKIEVQ 175  
 DB 403 ITGKXLMGKSVALLAQVLMFVGMTFMSNAHRSLAGMPRTAE---QVRN-EEEMAA 458  
 QY 176 SH---PANTAFCSLLQAQSL-----PRMAPQ--DSLPGED 211  
 DB 459 GSIGELNAQVILGILLFVSTLLFVLVVVTGLDGAEPGLPANVEAYDL-SGPED 514

RESULT 14  
 Q8WDS8 PRELIMINARY; PRT; 379 AA.  
 AC Q8WDS8;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Cytochrome b.  
 GN CYTB.  
 OS Lepidostiren paradoxus (South American lungfish).  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Dipnoi; Lepidostirentiformes; Lepidostirentidae; Lepidostiren.  
 OX NCBI\_TaxID=7883;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Denk A.G., Brinkmann H., Zitzler J., Joss J.U., Meyer A.;  
 RT "Complete Mitochondrial Genome Sequences of South American (Lepidostiren paradoxus) and Australian (Neoceratodus forsteri) lungfishes: Sarcopterygian - Actinopterygian division confirmed."  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF302934; AAL55248.1; -;  
 DR InterPro: IPR00179; Cyt\_b\_b6.  
 DR Pfam: PF00032; cytochrome\_b\_c\_1.  
 DR Pfam: PF00033; cytochrome\_b\_N\_1.

DR PROSITE: PS00192; CYTOCHROME B HEME; UNKNOWN 1.  
 DR PROSITE: PS00193; CYTOCHROME\_B\_OO; UNKNOWN 1.  
 KW Mitochondrion  
 SQ SEQUENCE 379 AA; 42889 MW; 02118DA863322CCB CRC64;

Query Match 5.4%; Score 91; DB 8; Length 379;  
 Best Local Similarity 28.1%; Pred. No. 4.6;  
 Matches 45; Conservative 20; Mismatches 51; Indels 44; Gaps 9;

QY 11 LILFLVMSLTVTRHLSGGSGGDFSNVLPPLPLPLAGLVQIIIFLGT 62  
 DB 230 LMGFTLLMLLIIISL-----FSPNLSDPENTPMPPLVTPPHIKPKWFLFAYA 280  
 QY 63 -----GTTALAF---LVMPVLH-GRNLLPFRSLSSWPFMLTALAVLONM 106  
 DB 281 ILRSIPNKLGVIALAFSILILPLFPLHOKKQRTAVRPI-SQLMFLLVSTMLILT-- 337  
 QY 107 AAHVFLETHGHQPOLTRRVLYAATFLF---PLVVV 142  
 DB 338 ---WICQVPE-HPIYIGQIASTYFMLPLALPLTAL 373

RESULT 15  
 Q37632 PRELIMINARY; PRT; 229 AA.  
 AC Q37632;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1) (fragment).  
 GN COII.  
 OS Rnagolepis alternata.  
 OC Mitochondrion.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Tephritidae; Tephritidae; Rnagolepis.  
 OX NCBI\_TaxID=47154;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97159559; PubMed=9007018;  
 RA Smith J.J., Bush G.L.;  
 RT "Phylogeny of the genus Rnagolepis (Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase II."  
 RL Mol. Phylogenet. Evol. 7:33-43(1997).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 2 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICCYTOCHROME C + 2 H(2)O.  
 CC -1- COFACTOR: COPPER A AND HEME GROUP (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL: U53260; AAB50353.1; -;  
 DR InterPro: IPR001505; Copper\_CuA.  
 DR InterPro: IPR002429; Cyt\_c\_Ox\_2.  
 DR Pfam: PF00116; COX2; 1.  
 DR Pfam: PF02790; COX2\_TM; 1.  
 DR PRINTS: PR01166; CYCOXIDASEII.  
 DR ProDom: PD000131; Copper\_CuA; 1.  
 DR PROSITE: PS00078; COX2; 1.  
 KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;  
 KW Respiratory chain; Transmembrane.  
 FT NON\_TER 229  
 SQ SEQUENCE 229 AA; 26559 MW; 2F99E26A33BEC148 CRC64;

Query Match 5.3%; Score 90.5; DB 8; Length 229;  
 Best Local Similarity 22.0%; Pred. No. 3;  
 Matches 63; Conservative 39; Mismatches 81; Indels 103; Gaps 14;



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:26:57 ; Search time 10.5439 Seconds  
(without alignments) 1278.447 Million cell updates/sec

Title: US-09-816-653a-4  
Perfect score: 1695  
Sequence: 1 CYISALVLSCLTFLVLMRS.....APPTLMESQGFMRKLVG 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	94	5.5	476 1	TR16_HUMAN
2	93	5.5	2016 1	C1NS_HUMAN
3	91	5.4	346 1	RX_HUMAN
4	90.5	5.3	625 1	N1FA_AZOB
5	89.5	5.3	395 1	YJGN_SALTY
6	89	5.3	3164 1	TEGU_HSV1
7	88.5	5.2	107 1	YGBE_ECOLI
8	87	5.1	621 1	GRBA_MOUSE
9	86	5.1	614 1	YDNK_IACIC
10	85.5	5.0	446 1	NYSR_PIG
11	84.5	5.0	736 1	DV12_HUMAN
12	84.5	5.0	1508 1	A10A_MOUSE
13	84	5.0	193 1	PMP2_MOUSE
14	84	5.0	321 1	NUIM_DINSE
15	84	5.0	527 1	SPUB_SERMA
16	83	4.9	222 1	YJ73_SYNY3
17	83	4.9	1840 1	C1N4_RAT
18	82.5	4.9	228 1	COX2_ZOON
19	82.5	4.9	300 1	Y223_TAEIN
20	82	4.8	347 1	NU2M_PANTR
21	81	4.8	335 1	YEB1_MYCAV
22	81	4.8	637 1	LBR_CHICK
23	81	4.8	1499 1	A10C_HUMAN
24	80.5	4.7	1038 1	ATA3_MOUSE
25	80	4.7	329 1	YCLC_PSEPU
26	80	4.7	379 1	CYB_ARPEI
27	80	4.7	654 1	SCS5_YEAST
28	80	4.7	859 1	SPH2_HUMAN
29	80	4.7	888 1	M3KC_HUMAN
30	80	4.7	944 1	MMU3_MYCTU
31	79.5	4.7	353 1	IL8B_MACMU
32	79.5	4.7	739 1	PURL_CAUCR
33	79.5	4.7		

34	79	4.7	235 1	CDBA_HUMAN	P01732 homo sapien
35	79	4.7	1304 1	S3B1_HUMAN	O75533 homo sapien
36	79	4.7	1304 1	S3B1_MOUSE	O99b9 mus musculu
37	79	4.7	1307 1	S3B1_XENLA	O57683 xenopus lae
38	78.5	4.6	511 1	WVIN_BUCAL	P57415 buchnera ap
39	78.5	4.6	721 1	BNP1_TORCA	P14400 corpeo cal
40	78	4.6	347 1	NU2M_BALPH	P24970 balaeopter
41	78	4.6	351 1	KLZ2_RAT	O9c9b8 ratius norv
42	78	4.6	1473 1	NAU1_HUMAN	O9c9b0 homo sapien
43	77.5	4.6	342 1	HUPK_AZOVI	P31878 azobacter
44	77.5	4.6	342 1	RX_MOUSE	O35602 mus musculu
45	77.5	4.6	348 1	HOXV_AZOVI	P40597 azobacter

## ALIGNMENTS

## RESULT 1

TR16\_HUMAN STANDARD: PRT: 476 AA.

AC 015654; 015275; 015170; 09BTB2; 09ONT4;

DT 01-NOV-1997 (Rel. 35, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Thyroid receptor interacting protein 6 (TRIP6) (OPA-interacting

DE protein 1) (Zyxin related protein 1) (ZRP-1).

GN TRIP6 OR OIPI

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN 1

RP SEQUENCE FROM N.A.

RX MEDLINE=98260683; PubMed=9598321;

RA Y. J., Beckerle M.C.;

RT "The human TRIP6 gene encodes a LIM domain protein and maps to

chromosome 4p16.3, a region associated with tumorigenesis.";

RL Genomics 49:314-316(1998).

RN 2

RP SEQUENCE FROM N.A.

RX MEDLINE=99329089; PubMed=10400701;

RA Murthy K.K., Clark K., Fortin Y., Shen S.-H., Banville D.;

RT "ZRP-1, a zyxin-related protein, interacts with the second PDZ domain

of the cytosolic protein tyrosine phosphatase hPTP.";

RL J. Biol. Chem. 274:20679-20687(1999).

RN 3

RP SEQUENCE FROM N.A. AND VARIANT PHE-296.

RA TISSUE=Kidney, and Pancreas;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

RN 4

RP SEQUENCE OF 310-476 FROM N.A.

RX MEDLINE=95295757; PubMed=7776974;

RA Lee J.W., Choi H.-S., Gyuris U., Brent R., Moore D.D.;

RT "Two classes of proteins dependent on either the presence or absence

of thyroid hormone for interaction with the thyroid hormone

receptor.";

RL Mol. Endocrinol. 9:243-254(1995).

-1- SUBUNIT: Specifically interacts with the ligand binding domain of

the thyroid receptor (TR). Requires the presence of thyroid

hormone for its interaction. Interacts with PTPN13.

-1- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN KIDNEY, LIVER AND

LUNG. LOWER LEVELS IN HEART, PLACENTA AND PANCREAS.

-1- SIMILARITY: CONTAINS 3 LIM DOMAINS. THE LIM DOMAIN BINDS 2

ZINC IONS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC  EMBL: AJ001902; CAA05080.1; -
DR  EMBL: AF000974; AAB62222.1; -
DR  EMBL: AF038356; AAB62222.1; -
DR  EMBL: AF038356; AAD03037.1; JOINED.
DR  EMBL: AF038356; AAD03037.1; JOINED.
DR  EMBL: BC004249; AAH04249.1; -
DR  EMBL: BC004999; AAH04999.1; -
DR  EMBL: BC021540; AAH21540.1; -
DR  EMBL: LA0374; AAC41740.1; -
DR  GenBank: HGNC:12311; TRIP6.
DR  MIM: 602933; -; TRIP6.
DR  InterPro: IPR001781; LIM.
DR  Pfam: PF00412; LIM; 3.
DR  ProDom: PD000094; LIM; 3.
DR  SMART: SM00132; LIM; 3.
DR  PROSITE: PS00478; LIM DOMAIN 1; 2.
DR  PROSITE: PS50023; LIM DOMAIN 2; 3.
DR  LIM domain; Repeat; Metal-binding; Zinc; Polymorphism.
KW  DOMAIN 279 316 LIM 1.
FT  DOMAIN 339 398 LIM 2.
FT  VARIANT 399 467 LIM 3.
FT  VARIANT 296 296 LIM 3.
FT  CONFLICT 102 102 /FtId=VAR_013309.
FT  CONFLICT 106 106 S -> T (IN REF. 2; AAB62222).
FT  CONFLICT 135 135 E -> K (IN REF. 2; AAB62222).
FT  CONFLICT 310 312 S -> C (IN REF. 1).
FT  CONFLICT 461 476 CRAQ -> MFCF (IN REF. 4).
FT  SEQUENCE 476 AA: 50288 MW: 28470730A8FD CRC64:
      KVMRIQESAVITVDC -> RFGASRSSQPSGLTAESMK
      YALLSQFQPSFD (IN REF. 4).
      ZBA7C747DF30A8FD CRC64:

Query Match 5.5%; Score 94; DB 1; Length 476;
Best Local Similarity 27.4%; Pred. No. 2.8;
Matches 43; Conservative 14; Mismatches 66; Indels 34; Gaps 7;

OY 175 GSHAMTAFCSL-----LLQASLPRMAAPD--SIRPEDEEGM 214
DB 39 OPHRVA-FCEPSEOCYAPGCEPBERGAWGSHGVLOHOGCPADGGLRPSLDAEI 97
OY 215 QLLOTKDSMAKGRPCAXRGARWGLAYTLNPTLQVFRKTLGANGAPCCSLP--- 271
DB 98 DLLESTLAELNGRGHGRPRDQA-----YEPPEPAYRTGSLKRPASPLPASYGG 151
OY 272 GSPSITPAMOPAGPPHKGXVVCULMEPPXGSAAP 308
DB 152 PTPASVTTASTPACP---APVQVQVA-OPVRGCGP 184

RESULT 2
CINS HUMAN STANDARD; PRT; 2016 AA.
AC 014524;
DT 15-DEC-1998 (Rel. 37, Last Sequence Update)
DT 15-DEC-1998 (Rel. 37, Last Sequence Update)
DT 15-OCT-2001 (Rel. 40, Last Annotation Update)
DE Sodium channel protein, cardiac muscle alpha-subunit (HHL).
GN SCN5A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=92115699; PubMed=1309946;
RA Geleens M.E., George A.L. Jr., Chen L.Q., Chahine M., Horn R.,
RA Barclay R.L., Kallen R.G.;
RT "Primary structure and functional expression of the human cardiac
RT leucodioxin-insensitive voltage-dependent sodium channel.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:554-558(1992).
RN [2];
RA VARIANTS LQ73.

```

```

RX MEDLINE=95196273; PubMed=7889574;
RA Wang Q., Shen J., Splawski I., Atkinson D., Li Z., Robinson J.L.,
RA Moss A.J., Towbin J.A., Keating M.T.;
RT "SCN5A mutations associated with an inherited cardiac arrhythmia, long
RT QT syndrome.";
RL Cell 80:805-811(1995).
RN [3];
RP VARIANTS LQ73.
RX MEDLINE=96081224; PubMed=8541846;
RA Wang Q., Shen J., Li Z., Timothy K., Vincent G.M., Priori S.G.,
RA Schwartz P.J., Keating M.T.;
RT "Cardiac sodium channel mutations in patients with long QT syndrome,
RT an inherited cardiac arrhythmia.";
RL Hum. Mol. Genet. 4:1603-1607(1995).
RN [4];
RP VARIANT LQ73 1505-LYS->GLN-1507 DEL.
RX MEDLINE=95379949; PubMed=7651517;
RA Bennett P.B., Yazawa K., Makita N., George A.L. Jr.;
RT "Molecular mechanism for an inherited cardiac arrhythmia.";
RL Nature 376:683-685(1995).
RN [5];
RP VARIANT LQ73 GLY-1790.
RX MEDLINE=98349542; PubMed=9686753;
RA An R.H., Wang X.L., Kelem B., Benhorin J., Medina A., Goldmit M.,
RA Kase R.S.;
RT "Novel LQ73 mutation affects Na+ channel activity through
RT interactions between alpha- and beta1-subunits.";
RL Circ. Res. 83:141-146(1998).
RN [6];
RP VARIANT LQ73 GLN-1623.
RX MEDLINE=98165676; PubMed=9506831;
RA Makita N., Shirai N., Nagashima M., Matsuoaka R., Yamada Y., Tohne N.,
RA Kitabatake A.;
RT "A de novo missense mutation of human cardiac Na(+) channel exhibiting
RT novel molecular mechanisms of long QT syndrome.";
RL FEBS Lett. 423:5-9(1998).
RN [7];
RP VARIANT LQ73 GLY-1839.
RA Benhorin J., Goldmit M., Maccluer J.W., Blangero J., Goffen R.,
RA Leibovitch A., Rahat A., Wang Q., Medina A., Towbin J., Kelem B.;
RT "Identification of a new SCN5A mutation, D1840G, associated with the
RT long QT syndrome.";
RL Hum. Mutat. 12:72-72(1998).
RN [8];
RP VARIANT LQ73 GLN-1623.
RA Yamagishi H., Furutani M., Kamisago M., Morikawa Y., Kojima Y.,
RA Hino Y., Furutani Y., Kimura M., Imamura S.-I., Takao A., Momma K.,
RA Matsuoaka R.;
RT "A de novo missense mutation (R1623G) of the SCN5A gene in a Japanese
RT girl with sporadic long QT syndrome.";
RL Hum. Mutat. 12:481-481(1998).
RN [9];
RP VARIANTS LQ73 MET-1304 AND MET-1645, AND VARIANT ASN-1500.
RX MEDLINE=99439526; PubMed=10508990;
RA Wattanachitichagoon D., Vesely M.R., Duggan P., Levine J.C.,
RA Blume E.D., Wolff G.S., Edwards S.B., Beggs A.H.;
RT "Sodium channel abnormalities are infrequent in patients with long QT
RT syndrome: identification of two novel SCN5A mutations.";
RL Am. J. Med. Genet. 86:470-476(1999).
RN [10];
RP VARIANT LQ73 LYS-1784.
RX MEDLINE=99307063; PubMed=10377081;
RA Wei J., Wang D.W., Allings M., Fish F., Wathen M., Roden D.M.,
RA George A.L. Jr.;
RT "Congenital long-QT syndrome caused by a novel mutation in a conserved
RT acidic domain of the cardiac Na+ channel.";
RL Circulation 99:3165-3171(1999).
RN [11];
RP VARIANTS LQ73 N-1114; V-1501; L-1623; H-1644 AND N-1787.
RX MEDLINE=20432616; PubMed=10973849;
RA Splawski I., Shen J., Timothy K.W., Lehmann M.H., Priori S.,
RA Robinson J.L., Moss A.J., Schwartz P.J., Towbin J.A., Vincent G.M.,
RA Keating M.T.;

```



RT Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNEL, and KCNE2."

RL Circulation 102:1178-1185(2000).

CC -1- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED CONFORMATION IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH WHICH NA+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT. IT IS A TETRODOTOXIN-RESISTANT NA+ CHANNEL ISOPFORM. THIS CHANNEL IS RESPONSIBLE FOR THE INITIAL UPSTROKE OF THE ACTION POTENTIAL IN THE ELECTROCARDIOGRAM.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN HUMAN ATRIAL AND VENTRICULAR CARDIAC MUSCLE BUT NOT IN ADULT SKELETAL MUSCLE, BRAIN, MYOMETRIUM, LIVER, OR SPLEEN.

CC -1- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5 HYDROPHOBIC SEGMENTS (S1,S2,S3,S5,S6) AND ONE POSITIVELY CHARGED SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.

CC -1- DISEASE: DEFECTS IN SCN5A ARE THE CAUSE OF LONG QT SYNDROME TYPE 3 (LQT3). AN AUTOSOMAL DOMINANT CARDIAC DISEASE CHARACTERIZED BY RECURRENT SYNCOPE AND SUDDEN CARDIAC DEATH.

CC -1- MISCELLANEOUS: NA+ CHANNELS IN MAMMALIAN CARDIAC MEMBRANE HAVE FUNCTIONAL PROPERTIES QUITE DISTINCT FROM NA+ CHANNELS IN NERVE AND SKELETAL MUSCLE.

CC -1- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.

CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC -1- DATABASE: NAME=LQTSdb; NOTE=SCN5A mutations page; WWW="http://www.ssi.dk/en/forekting/lqtsdb/scns.htm".

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M77235; AAA5644.1; -

DR Genbank: M77235; AAA5644.1; -

DR HMM: 600163; -

DR InterPro: IPR001682; Ca/Na\_Pore.

DR InterPro: IPR002111; Cal\_channel\_Tppl.

DR InterPro: IPR000048; IQ\_region.

DR InterPro: IPR000636; M+channel\_nlg.

DR InterPro: IPR001696; Na\_channel.

DR Pfam: PR00520; ion\_trans\_4.

DR Pfam: PR00612; IQ\_1.

DR PRINTS: PR00170; NACCHANNEL.

DR SMART: SM00015; IQ\_1.

DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel; Glycoprotein; Repeat; Multigene family; Phosphorylation; Polymorphism; Disease mutation; Long QT syndrome.

KW TRANSMEM 127 150 S1 OF REPEAT I (POTENTIAL).

FT TRANSMEM 159 178 S2 OF REPEAT I (POTENTIAL).

FT TRANSMEM 192 210 S3 OF REPEAT I (POTENTIAL).

FT TRANSMEM 217 236 S4 OF REPEAT I (POTENTIAL).

FT TRANSMEM 253 276 S5 OF REPEAT I (POTENTIAL).

FT TRANSMEM 290 315 S6 OF REPEAT I (POTENTIAL).

FT TRANSMEM 330 355 S1 OF REPEAT II (POTENTIAL).

FT TRANSMEM 371 396 S2 OF REPEAT II (POTENTIAL).

FT TRANSMEM 412 437 S3 OF REPEAT II (POTENTIAL).

FT TRANSMEM 448 473 S4 OF REPEAT II (POTENTIAL).

FT TRANSMEM 489 514 S5 OF REPEAT II (POTENTIAL).

FT TRANSMEM 526 551 S6 OF REPEAT II (POTENTIAL).

FT TRANSMEM 563 588 S1 OF REPEAT III (POTENTIAL).

FT TRANSMEM 600 625 S2 OF REPEAT III (POTENTIAL).

FT TRANSMEM 637 662 S3 OF REPEAT III (POTENTIAL).

FT TRANSMEM 674 699 S4 OF REPEAT III (POTENTIAL).

FT TRANSMEM 711 736 S5 OF REPEAT III (POTENTIAL).

FT TRANSMEM 748 773 S6 OF REPEAT III (POTENTIAL).

FT TRANSMEM 785 810 S1 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 822 847 S2 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 859 884 S3 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 896 921 S4 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 933 958 S5 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 965 990 S6 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 1002 1027 S1 OF REPEAT V (POTENTIAL).

FT TRANSMEM 1039 1064 S2 OF REPEAT V (POTENTIAL).

FT TRANSMEM 1071 1096 S3 OF REPEAT V (POTENTIAL).

FT TRANSMEM 1108 1133 S4 OF REPEAT V (POTENTIAL).

FT TRANSMEM 1145 1170 S5 OF REPEAT V (POTENTIAL).

FT TRANSMEM 1182 1207 S6 OF REPEAT V (POTENTIAL).

FT TRANSMEM 1214 1239 S1 OF REPEAT VI (POTENTIAL).

FT TRANSMEM 1251 1276 S2 OF REPEAT VI (POTENTIAL).

FT TRANSMEM 1283 1308 S3 OF REPEAT VI (POTENTIAL).

FT TRANSMEM 1315 1340 S4 OF REPEAT VI (POTENTIAL).

FT TRANSMEM 1352 1377 S5 OF REPEAT VI (POTENTIAL).

FT TRANSMEM 1389 1414 S6 OF REPEAT VI (POTENTIAL).

FT TRANSMEM 1426 1451 S1 OF REPEAT VII (POTENTIAL).

FT TRANSMEM 1458 1483 S2 OF REPEAT VII (POTENTIAL).

FT TRANSMEM 1490 1515 S3 OF REPEAT VII (POTENTIAL).

FT TRANSMEM 1522 1547 S4 OF REPEAT VII (POTENTIAL).

FT TRANSMEM 1559 1582 S2 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 1589 1612 S3 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 1623 1644 S4 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 1660 1682 S5 OF REPEAT IV (POTENTIAL).

FT TRANSMEM 1748 1772 S6 OF REPEAT IV (POTENTIAL).

FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 740 740 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 841 841 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1365 1365 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1374 1374 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1380 1380 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1388 1388 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1736 1736 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1774 1774 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1955 1955 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARIANT 558 558 H -> R.

FT VARIANT 1114 1114 /FTID=VAR 008955.

FT VARIANT 1304 1304 D -> N (IN LQT3).

FT VARIANT 1304 1304 T -> M (IN LQT3).

FT VARIANT 1325 1325 /FTID=VAR 008956.

FT VARIANT 1325 1325 N -> S (IN LQT3).

Query Match 5.5%; Score 93; DB 1; Length 2016;

Best Local Similarity 23.8%; Pred. No. 16;

Matches 67; Conservative 31; Mismatches 113; Indels 70; Gaps 13;

QY 35 QRSNLFVSVPLEPLPLAGLVYQIIFLGTALAFVLPVHGRLLPFRLESSEMPFW 94

DB 778 QQGMNIFD-----SIVILSMLGLSRMSNL--SVLRSEFLVFPKLAKSWPTL 825

QY 95 -----LTLALAVILONMAAHV-----FLETHDGPOLNTR---RVLY 129

DB 826 NTLIKIKIGNSVGLNLVLALIVLPIFAVGMQLGKQVSELRDSDGLPRHMMWDF 885

QY 130 AATFLPLPLVIVGAMKXKXSPSIALNHPYGYTTRNPLKIEVSQSHRAMPAFCSILQ 189

DB 886 HAFLIIF--RIICGEMIEIEMDCMEVSGQS---LCLLVFLVWVIGNLVVLVFLALLLS 940

QY 190 AOSLPRTMAAPQDSLRGCEDEGQQLQTKDSMAKAGARGRARWGLAYTTL--HNP 248

DB 941 SFS-----ADNLTAPDDBDRNNNLQ--AAARIQRLARVVKTTDFCCGLRRHP 989

QY 249 TLQVFRKTLALGANGAP--CSSLPQSPSPSTPAMPQAPGPD 288

DB 990 -----QKPALAAQGLPSCIAATPVSP-----PPE 1015

RESULT 3

RX\_HUMAN STANDARD; PRT: 346 AA.

AC Q9Y2V3;

DT 16-OCT-2001 (Rel. 40; Created)

DT 16-OCT-2001 (Rel. 40; Last sequence update)

DT 15-JUN-2002 (Rel. 41; Last annotation update)

DE Retinal homeobox protein Rx (Retina and anterior neural fold homeobox protein).

GN RAY OR RX.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

SEQUENCE FROM N.A.  
 MEDLINE=2009286; PubMed=10625658;  
 Kimura A., Singh D., Wawrousek E.F., Kikuchi M., Nakamura M.,  
 Shinohara T.,  
 "Both PCE-1/RX and OTX/CRX interactions are necessary for  
 photoreceptor-specific gene expression."  
 J. Biol. Chem. 275:1152-1160(2000)  
 - FUNCTION: plays a critical role in eye formation by regulating the  
 initial specification of retinal cells and/or their subsequent  
 proliferation. binds to the photoreceptor conserved element-1  
 (PCE-1/Ret 1) in the photoreceptor cell-specific arrestin  
 promoter.  
 - SUBCELLULAR LOCATION: Nuclear.  
 - TISSUE SPECIFICITY: Expressed in the developing eye and weakly  
 expressed in the adult retina.  
 - SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 "HICOID" SUBFAMILY.  
 - SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 - SIMILARITY: CONTAINS 1 OAR DOMAIN.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: AF115392; AAD23438.1; -  
 HSSP: P06601; 1FJL.  
 DR Genew; HGNC:18662; RAX.  
 DR MIM; 601881; -  
 DR InterPro; IPR003654; Homeo OAR.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox\_1.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 DR PROSITE; PS50803; OAR; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 33 40 OCTAPEPTIDE MOTIF.  
 FT DNA BIND 136 195 HOMEBOX.  
 FT DOMAIN 323 336 OAR.  
 FT DOMAIN 329 333 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 276 282 POLY-PRO.  
 SQ SEQUENCE 346 AA; 36804 MW; 5F1D0512DCB90282 CRC64;  
 -----  
 Query Match 5.4%; Score 91; DB 1; Length 346;  
 Best Local Similarity 23.8%; Pred. No. 3.4;  
 Matches 48; Conservative 16; Mismatches 80; Indels 58; Gaps 6;

ID NIFA\_AZOBZ STANDARD; PRT; 625 AA.  
 AC P30667;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Nif-specific regulatory protein.  
 GN NIFA.  
 OS Azospirillum brasilense.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;  
 CC Azospirillum.  
 NX NCBI\_Taxid=192;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SP7.  
 RX MEDLINE=92140038; PubMed=1779763;  
 RA Liang Y.Y., Kaminaki P.A., Elmerich C.;  
 RT "Identification of a nifA-like regulatory gene of Azospirillum  
 brasilense Sp7 expressed under conditions of nitrogen fixation and in  
 the presence of air and ammonia."  
 RL Mol. Microbiol. 5:2735-2744 (1991).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=9312252; PubMed=1362170.  
 RA Liang Y.Y., de Zamaroczy M., Arsene F., Paquelin A., Elmerich C.;  
 RT "Regulation of nitrogen fixation in Azospirillum brasilense Sp7:  
 involvement of nifA, glnA and glnB gene products."  
 RL FEMS Microbiol. Lett. 79:113-119 (1992).  
 CC - FUNCTION: NIFA, A TRANSCRIPTIONAL ACTIVATOR, IS REQUIRED FOR  
 ACTIVATION OF MOST NIF OPERONS, WHICH ARE DIRECTLY INVOLVED IN  
 NITROGEN FIXATION. NIFA INTERACTS WITH SIGMA-54.  
 CC - SIMILARITY: THE CENTRAL REGION CONTAINS A SIGMA-54 FACTOR  
 INTERACTION ATP-BINDING DOMAIN.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: X60714; CAA43126.1; -  
 DR EIR; S18420; S18420.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003018; GAF.  
 DR InterPro; IPR002197; HTH\_Fis.  
 DR InterPro; IPR002078; Sig54\_interact.  
 DR Pfam; PF00158; Sigma54\_activat; 1.  
 DR Pfam; PF01590; GAF; 1.  
 DR Pfam; PF02954; HTH\_8; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR SMART; SM00065; GAF; 1.  
 DR TIGRfam; TIGR01199; HTH\_Fis; 1.  
 DR PROSITE; PS00675; SIGMA54\_INTERACT\_1; 1.  
 DR PROSITE; PS00676; SIGMA54\_INTERACT\_2; 1.  
 DR PROSITE; PS00688; SIGMA54\_INTERACT\_3; 1.  
 DR PROSITE; PS50045; SIGMA54\_INTERACT\_4; 1.  
 KW Nitrogen fixation; Transcription regulation; Activator;  
 FT ATP-binding; DNA-binding.  
 FT DOMAIN 205 433 SIGMA-54 FACTOR INTERACTION (POTENTIAL).  
 FT DOMAIN 434 582 INTER-DOMAIN LINKER.  
 FT DOMAIN 583 625 C-TERMINAL DNA-BINDING DOMAIN.  
 FT NP BIND 233 240 ATP (POTENTIAL).  
 FT NP BIND 296 305 ATP (POTENTIAL).  
 FT METAL 447 447 BY SIMILARITY.  
 FT METAL 452 452 BY SIMILARITY.  
 FT DNA BIND 597 616 H-T-H MOTIF (BY SIMILARITY).  
 SQ SEQUENCE 625 AA; 67855 MW; D060EAB6CB1FCC3 CRC64;  
 -----  
 Query Match 5.3%; Score 90.5; DB 1; Length 625;  
 Best Local Similarity 27.7%; Pred. No. 7;  
 Matches 48; Conservative 9; Mismatches 55; Indels 61; Gaps 9;

RESULT 4  
 NIFA\_AZOBZ

```

OY 184 CSLLLOAOAIL--PRMTAPODSLRPGHEDEGMQLQTKDSMAKGA-----RPG-----230
DB 447 CSINLCNSSVLFQYRTTLCGASVGLAP-----SMQPGAINRPPRGCGDPA 492
OY 231 -----AXRGARGLAYTLTHNPLOVRKTKALLGANG-----NOP 266
DB 493 ANAKPTAMPAPVPEPDAGAAARGRP-----ARRVPPPLAGLRRA--GSGSPDPDPACP 546
OY 267 CSSLPGSPSITPAMOPAGPPDHGXGVCLHWPXGXGAPPTWLMSQCGFW 319
DB 547 CFSRAPLPPOAPPSPAAPPP-----AAEVPLD-EPESGSLRDLRLWAMERTGM 595

RESULT 5
YUGN_SALTY STANDARD; PRT; 395 AA.
ID YUGN_SALTY
AC 008022;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yj9N (ORF X).
GN YUGN OR STM4474.
OS Salmoneilla typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L12 / SGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11671609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du R., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmoneilla enterica serovar Typhimurium
RT L12."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 1-315 FROM N.A.
RC STRAIN=L12;
RX MEDLINE=94075212; PubMed=8253666;
RA Persson B.C., Bjorck G.R.;
RA "Isolation of the gene (mlaE) encoding the hydroxylase involved in
RT the synthesis of 2-methylthio-cis-ribozeatin in RNA of Salmoneilla
RL typhimurium and characterization of mutants."
J. Bacteriol. 175:7776-7785(1993).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1-
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008910; AL23293.1; -
DR EMBL; X73368; CA51785.1; -
DR PIR; S34364; S34364.
DR StyGene; SG10443; yj9N.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 230 250 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT TRANSMEM 312 313 SI -> R (IN REF. 2).
FT CONFLICT 395 AA; 44107 MW; 04C59F2938B731E2 CRC64;
SQ SEQUENCE

```

```

Query Match 5.3%; Score 89.5; DB 1; Length 395;
Best Local Similarity 26.2%; Pred. No. 5.1;
Matches 37; Conservative 20; Mismatches 45; Indels 39; Gaps 6;

OY 5 ALVLSCLTFLVLMRSLVTHRLGSGGSGDGFSSWNVFSPVPLPLPLAGLVQOITFFLGT 64
DB 73 ALFISLVLFVIVISLIEHGHPLG-----FTLF-----GLLI-ALIPPMV 116
OY 65 TALAIVLIMPTLHGRNLLFRSLSSMPFMTLALAVIIONMAHWVLETHDHPQLTN 124
DB 117 KGLQYQAMMTSLNQHVFQCSNRRAW--WYFALPVL--MVA----- 156
OY 125 RRVLVAATFLPLPLVLVGAM 145
DB 157 ---LYIVLYISLVIAVGL 174

RESULT 6
TEGU_HSV11 STANDARD; PRT; 3164 AA.
ID TEGU_HSV11
AC P10220;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein (Viron protein UL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeech D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RA "The complete DNA sequence of the long unique region in the genome of
RT Herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1988).
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC HSV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14112; CA32311.1; -
DR PIR; I30085; WMBH6.
DR InterPro; IPR005210; Herpes_UL36.
DR Pfam; PF03586; Herpes_UL36_1.
KW Repeat.
FT DOMAIN 2911 2980 35 X 2 AA TANDDEM REPEATS OF P-Q.
FT DOMAIN 3164 AA; 335857 MW; C5SD31FFAF9FE3F4 CRC64;
SQ SEQUENCE

Query Match 5.3%; Score 89; DB 1; Length 3164;
Best Local Similarity 36.9%; Pred. No. 52;
Matches 24; Conservative 4; Mismatches 17; Indels 20; Gaps 3;

OY 247 NPTLOVRKTKALLGANGAOPCSLPGSPSITPAMOPAGPPDHGXGVCLHWPXGXGA 306
DB 2824 DPTAPV-----LGRNPAFTSSSPAGSPSPVAVOPVAPF-----PISG-P 2863
OY 307 PPTWL 311
DB 2864 PPTVL 2868

RESULT 7
YGBE_ECOLI

```

```

ID      YGBE_ECOLI     STANDARD;          PRT;          107 AA.
DT      P46141; Q46895;
DT      01-NOV-1995 (Rel.13, Created)
DT      01-NOV-1997 (Rel.35, Last sequence update)
DT      16-OCT-2001 (Rel.40, Last annotation update)
DE      Hypothetical protein ygbE.
GN      YGBE OR B2749.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      (1)
RP      SEQUENCE FROM N.A.
RP      STRAIN=K12 / MG1655.
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Greer J., Davis N.M., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      (2)
RP      SEQUENCE OF 1-81 FROM N.A.
RP      STRAIN=K12;
RX      MEDLINE=92268080; PubMed=1316900;
RA      Lehy T.S., Vogt T.F., Suo Y.;
RT      "The DNA sequence of the sulfate activation locus from Escherichia
RT      coli K-12.";
RL      J. Biol. Chem. 267:10405-10410(1992).
RN      (3)
RP      IDENTIFICATION.
RX      MEDLINE=96032851; PubMed=7567469;
RX      Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RX      Danchin A.;
RA      "Detection of new genes in a bacterial genome using Markov models for
RT      three gene classes.";
RL      Nucleic Acids Res. 23:3554-3562(1995).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U29579; AAA69259.1; -
DR      EMBL; AE000358; AAC75971.1; -
DR      EMBL; M74586; -; NOT ANNOTATED_CDS.
DR      EcoGene; EG12707; ygbE.
DR      Hypothetical protein; Complete proteome.
FT      CONFLICT      67      81      VAVVVGIALHSLMDG -> SRSGGDCAAFIDRK (IN
FT      REF. 1).
FT      SEQUENCE      107 AA; 12017 MW; E3ED0F889EA9C21F CRC64;
SQ
Query Match          5.2%; Score 88.5; DB 1; Length 107;
Best Local Similarity 24.6%; Pred. No.1.5;
Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;
OY      33 DGGQSNLFSVPLPPLPLAGLVQOQIFLGTTLALAFVLMPLV-GRNLLFPRSLESSW 91
DB      18 DEETYS-----LPGAVV-----GRISWLFPLAMPMLIYSSNTLFP-PIYTM 58
OY      92 PFWTLTALAVIIONMAAHWFLETHDGHPOLNRRVLAATFLFLPLNLVGM 145
DB      59 PFFALMPVAVVVGIALH-----SLMDGLRYSIVFTLVTVGIMFGAL 101

```

DT 15-JUL-1999 (Rel. 38, last sequence update)  
DE 16-OCT-2001 (Rel. 40, last annotation update)  
GN Growth factor receptor-bound protein 10 (GRB10 adaptor protein).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
ON NCBI\_TaxId=10090;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Swiss;  
RA MEDLINE=95249278; PubMed=7731717;  
RA Ooi J., Yajnik V., Immanuel D., Gordon M., Moskow J.J., Buchberg A.,  
RA Margolis B., of Grb10 reveals a new family of SH2 domain proteins";  
RT "The cloning of Grb10 reveals a new family of SH2 domain proteins";  
RL Oncogene 10:1621-1630(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97216049; PubMed=9062339;  
RA Lavoie L., Giorgio F., Chow J.C., Baquero J.A., Hansen H., Ooi J.,  
RA Zhu J., Riedel H., Smith R.J.;  
RT "The adapter protein grb10 associates preferentially with the insulin  
RT receptor as compared with the IGF-1 receptor in mouse fibroblasts.";  
RL Clin. Invest. 99:830-837(1997).  
CC -I- FUNCTION: PLAYS A FUNCTIONAL ROLE IN INSULIN AND IGF-1 SIGNALING.  
CC MAY SERVE TO POSITIVELY LINK THE INSULIN AND IGF-1 RECEPTORS TO AN  
CC UNCHARACTERIZED MITOGENIC SIGNALING PATHWAY. INTERACTS WITH THE  
CC CYTOSOLIC DOMAIN OF THE AUTOPHOSPHORYLATED INSULIN RECEPTOR  
CC WHICH IS THEN INHIBITED. THE INTERACTION IS MEDIATED BY THE SH2  
CC DOMAIN. ALSO BINDS ACTIVATED PLATELET-DERIVED GROWTH FACTOR  
CC RECEPTOR AND EPIDERMAL GROWTH FACTOR RECEPTOR.  
CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -I- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -I- SIMILARITY: BELONGS TO THE GRB7/10/14 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.iesb-sib.ch/announce/>  
CC or send an email to [license@iesb-sib.ch](mailto:license@iesb-sib.ch)).  
CC -----  
DR EMBL; U18996; AAB33687.1; -.  
DR EMBL; AF022072; AAB72103.1; -.  
DR HSSP; O60880; ID12.  
DR MGD; MGI:103232; Grb10.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR001159; RA domain.  
DR InterPro; IPR000980; SH2.  
DR Pfam; PR00017; SH2; 1.  
DR Pfam; PR00169; PH; 1.  
DR Pfam; PF00788; RA; 1.  
DR PRINTS; PR00401; SH2DOMAIN.  
DR ProDom; PD000093; SH2; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00314; RA; 1.  
DR SMART; SM00252; SH2; 1.  
DR PROSITE; PS50003; PH DOMAIN; 1.  
DR PROSITE; PS50001; SH2; 1.  
DR SH2 domain, Alternative splicing.  
FT DOMAIN 318 427 PH.  
FT DOMAIN 520 601 SH2.  
FT VARSPLIC 117 141 MISSING (IN ISOFORM 2).  
FT CONFLICT 491 492 NG -> KR (IN REF. 2).  
SQ SEQUENCE 621 AA; 70471 MW; 2A9A45D842468A7 CRC64;

```

OY 172 EVSOSHPRNTAFCSL-----LLO-----AQSLLPRTMAPODSLPEGEDEG 213
DB 4 DINSSVSEINSCNNQSDTTPALLEDOQHASNOGASSSRRQPSPOKQKQKQPHVI 63
OY 214 MOLLQTKDSMAKGRPGAXRGRARGLAYTLNHPITQVFRKATALLGANGAOPCSSLPSS 273
DB 64 LRLRLOEEDQQLMTA-----SLPAIPNPPE-----LPGA-----ADGS 96
OY 274 PPSITPAM---OPAGPP-DHXGVYVCLHHEPKXGSA-P-TWIMESQOGFMRKLL 323
DB 97 PPSVAPSSLPSPQSPAKHGCRCCKWIPGENTRGKRTKIMWOPPPQQLSKL 151

RESULT 9
YDNK_LACLC STANDARD; PRT; 614 AA.
AC P42377;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 70.0 kDa protein in daak 3'region (ORF4).
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=13359;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=MG1363;
RX MEDLINE=94172317; PubMed=8126443;
RA Eaton T.J., Sheatman C.A., Gasson M.J.;
RT Cloning and sequence analysis of the daak gene region of Lactococcus
RT lactis subsp. lactis.
RT J. Gen. Microbiol. 139:1253-1263(1993).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X76642; CAA54090.1; -
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 69869 MW; B1146136D0989D9 CRCK4;

Query Match 5.1%; Score 86; DB 1; Length 614;
Best Local Similarity 19.9%; Pred. No. 15;
Matches 49; Conservative 35; Mismatches 74; Indels 88; Gaps 11;

OY 5 ALVISCILTVYMRSLVTHRLSGSGSGDQSPWNLSEV-----PLPPLAGLIVQOI 58
DB 332 SLIFSCSLVLMQPLI-----FPMFLSFIFITVODPARFSTLGLF-- 374
OY 59 IFFIGTALAFVLMFVLRNLFFRSLSSSPFMTLALAVILQNMAMWVLETHDG 118
DB 375 -----SALSIVLILPIL-----LDKISGTSYVLTIGLIVFSIL--G 410
OY 119 HPOLTRN-----RVLYAA-----FELLPLNVLVGMXXXGSPSIAIRPPGYV 163
DB 411 FAERNRIQSGQPLPSASQSLNKTPEYIMENPDSIAIGYL---PQYIGHNDPYEK 466
OY 164 TYRNFLEIE--VSOSHPRNTAFCSLLLOAQSLLPRTMAPODSLPEGEDEGQLQTKD 221
DB 467 TIQGFYDKRVYGMKRNQAMT-----YLISQKGLPEGLA-----KSIQISD 506
OY 222 SMAKGA 227
DB 507 YSKKGS 512

```

```

AC 097969;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neuropeptide Y receptor type 5 (NPY-5) (NPY-Y5 receptor) (Y5
DE receptor).
GN NPY5R OR NPYR5.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1] SEQUENCE FROM N.A.
RP Wraith A., Tornsten A., Chardon P., Harbitz I., Chowdhary B.P.,
RA Anderson L., Larhammar D.;
RT "Porcine NPY receptors NPY1R, NPY2R and NPY5R: cloning, mapping and
RT comparative analysis.";
RT Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LMD; TISSUE=Kidney;
RA Ito Y., Minezawa M.;
RT "Sus scrofa gene for neuropeptide Y receptor type 5, complete cde.";
RT Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: RECEPTOR FOR NEUROPEPTIDE Y AND PEPTIDE YY. THE ACTIVITY
CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS THAT INHIBITS ADENYLATE
CC CYCLASE ACTIVITY. SEEMS TO BE ASSOCIATED WITH FOOD INTAKE. COULD
CC BE INVOLVED IN FEEDING DISORDERS (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC HIGHEST TO TACHYKININS RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF106083; AAD13778.1; -
DR EMBL: AB019185; BAA34055.1; -
DR Interpro: IPR000276; GPCR_Rhodopn.
DR Pfam: PF00001; 7tm.1; 1
DR PRINTS: PR00237; GPCRHOPOPSN
DR PROSITE: PS00237; G_PROTEIN_RECP_F1_1; FALSE_NEG.
DR PROSITE: PS50262; G_PROTEIN_RECP_F1_2; 1
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 39
FT TRANSMEM 41 62
FT DOMAIN 63 74
FT TRANSMEM 75 95
FT DOMAIN 96 115
FT TRANSMEM 116 137
FT DOMAIN 138 157
FT TRANSMEM 158 178
FT DOMAIN 179 210
FT TRANSMEM 211 232
FT DOMAIN 233 311
FT TRANSMEM 359 391
FT TRANSMEM 392 404
FT TRANSMEM 405 428
FT DOMAIN 429 446
FT CARBOHYD 10 10
FT CARBOHYD 17 17
FT DISULFID 114 198
FT LIPID 442 442
SQ SEQUENCE 446 AA; 50474 MW; B7F616C2394C6CA0 CRCK64;

Query Match 5.0%; Score 85.5; DB 1; Length 446;
Best Local Similarity 21.0%; Pred. No. 12;
Matches 69; Conservative 40; Mismatches 109; Indels 111; Gaps 16;

```

QY 58 IIFETTTALAFVLMVPHGRN-----LLFPRS-----LESSWP 92  
 DB 53 LLGFMGN-----LLILMAVMRKRQKTVNELIGNAFSDILVLCFSPFTLTVLLQWM 108  
 QY 93 F-----MLTALAVILQNMAMHWLETHDGHQDQTNREVLVAAT 132  
 DB 109 FGKVMCHIMPFLOCVTVLSTLLISTAIVRHYMHG-----PVSNMLTANHGY 157  
 QY 133 FLPLPLANTLVAMAMXXXSPSIAIRHPPTGYTT-----RNPFLKIEVQSHPA 179  
 DB 158 FLIATVMTL-----GLAICSPLPFPHSLVELQESFGSAMLRSRYLCEWSPDSY 207  
 QY 180 MTAFCSLLLOQSLLP-----RTMA---APQDSLRCGEDEGMOL-LQTKDSM 223  
 DB 208 RIAFTISLLVQVILPLVCLTVSHTSYCRISGCLSSQDSKL--EENEMILVTLQP----- 261  
 QY 224 AKGAPGAXRG-RARMGALATLHPTLOVFRKTALGANGQPC-SSLPGSPSTTPAM 281  
 DB 252 AKRSGFOAKLSHFKM--TSTIRRHRRRYSKKTACVRPAPGPALESREGPPGKXGSM 319  
 QY 282 QPAGPPDH-XGAVVCLHWEPPXGAP 308  
 DB 320 OSOPPSNKFMPGVPTCFEVKPENSDVP 348

## RESULT 11

ID DVL2 HUMAN STANDARD; PRT; 736 AA.  
 AC 014641;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Segment polarity protein dishevelled homolog DVL-2 (Dishevelled-2) (DSH homolog 2).  
 GN DVL2.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND PHOSPHORYLATION. MEDLINE=97336056; Pubmed=9192851;  
 RA Semenov M.V., Snyder M.;  
 RT "Human dishevelled genes constitute a DHR-containing multigene family.";  
 RL Genomics 42:302-310(1997).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE SIGNAL TRANSDUCTION PATHWAY MEDIATED BY MULTIPLE WNT GENES.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -1- SIMILARITY: BELONGS TO THE DSH FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 DEP DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AF060612; AAB65243.1; --  
 DR HSSP: Q12923; 3PDZ.  
 DR Genew; HGNC:3086; DVL2.  
 DR MIM; 603151; --  
 DR InterPro; IPR000591; DEP.  
 DR InterPro; IPR001158; DIX.  
 DR InterPro; IPR003351; Dishevelled.  
 DR InterPro; IPR001478; PDZ.  
 DR Pfam; PF00595; PDZ; 1.  
 DR Pfam; PF00610; DEP; 1.

DR Pfam; PF00778; DIX; 1.  
 DR Pfam; PF02377; Dishevelled; 1.  
 DR Prodom; PD003639; DIX; 1.  
 DR SMART; SM00021; DAX; 1.  
 DR SMART; SM00049; DEP; 1.  
 DR SMART; SM00228; PDZ; 1.  
 DR PROSITE; PS00186; DEP; 1.  
 DR PROSITE; PS0106; PDZ; 1.  
 DR PROSITE; PS0106; PDZ; 1.  
 DR Developmental protein; Phosphorylation.  
 FT DOMAIN 7 12 POLY-GLY.  
 FT DOMAIN 235 240 POLY-ARG.  
 FT DOMAIN 267 339 PDZ.  
 FT DOMAIN 433 507 DEP.  
 FT DOMAIN 666 694 POLY-PRO.  
 SQ SEQUENCE 736 AA; 78947 MW; 4B4D95B6C3E531B CRC64;

Query Match 5.0%; Score 84.5; DB 1; Length 736;  
 Best Local Similarity 28.8%; Pred. No. 25;  
 Matches 36; Conservative 9; Mismatches 37; Indels 43; Gaps 6;

QY 204 SLRPGEGEMOLQTKDS--MAKAPGAXRGARWGL-----AVTLHNPLOVF 253  
 DB 630 SLRGGEGAG-----TSDGPPRSGSTGAPNIRAPGLHPGPPGMALETPMVMVM 684  
 QY 254 RKTALLGANGAQCSSLRGSPPSTTPANQAPGPDHGXVGVCLHWEPPXGAPPTWME 313  
 DB 685 -----MPPPPVPVFPVAPGAPGAP-----VRDLGSPPE-LTA 716  
 QY 314 SQQGF 318  
 DB 717 SRQSF 721

## RESULT 12

ID A10A MOUSE STANDARD; PRT; 1508 AA.  
 AC 054827;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Potential phospholipid-transporting ATPase VA (EC 3.6.3.1).  
 GN ATP10A OR ATP5C.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Teratocarcinoma;  
 RX MEDLINE=20473714; Pubmed=11015572;  
 RA Halleck M.S., Lawler J.F., Jr., Blackshaw S., Gao L., Nagarajan P., Hacker C., Pyle S., Newman U.T., Nakanishi Y., Ando H., Weinstock D., Williamson P., Schlegel R.A.;  
 RT "Differential expression of putative transblayer amphipath transporters.";  
 RL Physiol. Genomics 1:139-150(1999).  
 CC [2]  
 CC SEQUENCE OF 16-435 FROM N.A.  
 RX MEDLINE=98217376; Pubmed=9548971;  
 RA Halleck M.S., Pradhan D., Blackman C.F., Berkes C., Williamson P.L., Schlegel R.A.;  
 RT "Multiple members of a third subfamily of P-type ATPases identified by genomic sequences and ESTs.";  
 RL Genome Res. 8:354-361(1998).  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: FOUND IN TESTIS. ALSO DETECTED IN PETAL TISSUES.  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2 ATPASES). SUBFAMILY IV.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announce/> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; AF011337; AAC02902.1; -  
EMBL; AF156549; AAC09447.1; -  
MGD; MGI:1330809; Atplga.  
InterPro; IPR001757; ATPase\_E1-E2.  
InterPro; IPR001454; Hlgnaase/hydrlase.  
Pfam; PF00702; Hydrolase; 1.  
PRINTS; PR00119; CATAPASE.  
DR PROSITE; PS00154; ATPase\_E1-E2; 1.  
Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
KW Multigene family.  
KW Domain 1 79  
FT TRANSMEM 80 101 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 102 107 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 108 129 POTENTIAL.  
FT DOMAIN 130 313 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 314 335 POTENTIAL.  
FT DOMAIN 336 366 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 367 388 POTENTIAL.  
FT DOMAIN 389 1101 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1102 1122 POTENTIAL.  
FT DOMAIN 1123 1134 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1135 1154 POTENTIAL.  
FT DOMAIN 1155 1184 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1185 1206 POTENTIAL.  
FT DOMAIN 1207 1213 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1214 1236 POTENTIAL.  
FT DOMAIN 1237 1242 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1243 1263 POTENTIAL.  
FT DOMAIN 1264 1281 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1282 1306 POTENTIAL.  
FT DOMAIN 1307 1508 CYTOPLASMIC (POTENTIAL).  
FT MOD RES 431 431 PHOSPHORYLATION (BY SIMILARITY).  
FT METAL 1045 1045 MAGNESIUM (BY SIMILARITY).  
FT METAL 1049 1049 MAGNESIUM (BY SIMILARITY).  
FT DOMAIN 17 23 POLY-ARG.  
FT DOMAIN 471 474 POLY-GLU.  
FT CONFLICT 16 22 WRPRRR -> KLAAAK (IN REF. 2).  
FT CONFLICT 435 435 T -> L (IN REF. 2).  
SQ SEQUENCE 1508 AA; 168699 MW; DC9A0D99AF7EBB9E CRC64;

Query Match 5.0%; Score 84.5; DB 1; Length 1508;  
Best Local Similarity 21.5%; Pred. No. 53;  
Matches 53; Conservative 26; Mismatches 80; Indels 87; Gaps 11;

59 IFFLGITLALFLVIMPYLHGRNLLFRSLSSWPFVLTALAVLQNMMAHWVLETHDG 118  
1214 VFTWGTPTVLTALFTFLH---LGLETWTWLTWMLAGF----- 1250  
119 HPGLTNRRLVLAFTLFLPLNLVGMXXXCSPSIAIRHPPTGYVYTNFLKIEVSQSH 178  
1251 -----STFLPFSVALIYNTSCATY-----PSNPVTWOTLL-----GDP 1286  
179 AMTAFCSLLLOAGSLPRT-MAAPQSLRGEDEGMOLLO-----TDSMAKXGAP 229  
1287 LFYLTG-LIADIALLPRLFPKALQSLPFTQLOLRQALAKVLPKSPDPRKFAQGP 1345  
230 GAKRGARWGLAVTLLNPTLOVFRKTALLGANGAPCSLLFGSPS-----ITPAM 281  
1346 G-----HSETLSERKT-WG-----PFTLRDCAQASQFTOOLTCSP 1383  
287 QPAGP 287  
1384 EASGP 1389

PM2 MOUSE STANDARD; PRT; 193 AA.  
P42925; 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 22 kDa peroxisomal membrane protein.  
GN PM2 OR PM22.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
OK NCBI  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Swiss Webster; TISSUE=Liver;  
RX MEDLINE=96028630; PubMed=7551822;  
RA Bryant D.D., Wilson G.N.;  
RT "differential evolution and expression of murine peroxisomal membrane protein genes".  
RT Biochem. Mol. Med. 55:22-30(1995).  
RL -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal.  
CC -1- SIMILARITY: BELONGS TO THE MPV17 / PM22 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announce/> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; L28835; AAA39957.1; -  
DR MGD; MGI:107487; Pmp2.  
KW Transmembrane; Peroxisome.  
FT INIT MET 0  
FT DOMAIN 1 29 BY SIMILARITY.  
FT TRANSMEM 30 50 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 51 73 POTENTIAL.  
FT TRANSMEM 74 94 PEROXISOMAL (POTENTIAL).  
FT DOMAIN 95 112 POTENTIAL.  
FT TRANSMEM 113 133 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 134 171 POTENTIAL.  
FT TRANSMEM 172 192 PEROXISOMAL (POTENTIAL).  
SQ SEQUENCE 193 AA; 22134 MW; 493A467470381E0C CRC64;

Query Match 5.0%; Score 84; DB 1; Length 193;  
Best Local Similarity 28.1%; Pred. No. 6.4;  
Matches 34; Conservative 18; Mismatches 47; Indels 22; Gaps 4;

9 SCLTFLVMSRLVTHRLGSGSGSDGQFSWNLPSVPLPLAGL-LVQOLIFFLGTTA 66  
62 SRLLEVSGLLRLVGLFVTGPLSHLYLFMEYSVP-PEVPMASVGRLLDLRLFFAPFL 120  
67 LAFVLMPLYLHGRNLL-----FRSLSSWPFV-----LTALAVLQNMMA 107  
121 LFFPMNLLEGKNSVVAVMRSGFWPALQNMWMMVTPLOFINIVYPLCRVLFANMA 180  
108 A 108  
181 A 181

RESULT 14  
NUM DINGE STANDARD; PRT; 321 AA.  
ID NUM DINGE  
AC 079546;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE NADH-ubiquinone oxidoreductase chain 1 (EC 1.6.5.3).  
GN MTND1 OR ND1.  
OS Dinosaur semicarinatus (Akamata) (Dinosaur semicarinatum).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidodactylus: Squamata; Scleroglossae; Serpentes; Colubroidea;  
 CC Colubridae; Colubrinae; Dinodon.  
 OX NCBI\_TaxId=56549;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA MEDLINE=96393589; PubMed=9725849;  
 RA Kumazawa Y., Ota H., Nishida M., Ozawa T.;  
 RT The complete nucleotide sequence of a snake (Dinodon semicarinatus)  
 RL mitochondrial genome with two identical control regions.;  
 CC GenBank 150,313-329 (1996).  
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL, AB008539; BAA33022.1; -  
 DR InterPro: IPR001694; Resp\_NADH\_dhl.  
 DR Pfam: PF00146; NADHdh; 1.  
 DR PROSITE: PS00667; COMPLEX1\_ND1\_1; 1.  
 DR PROSITE: PS00668; COMPLEX1\_ND1\_2; 1.  
 DR Oxidoreductase; NAD: Ubiquinone, Mitochondrion; Transmembrane.  
 KW SEQUENCE 321 AA; 36003 MW; 6907ECBAAC2D9D CRC64;  
 SQ  
 Query Match 5.0%; Score 84; DB 1; Length 321;  
 Best Local Similarity 20.4%; Pred. No. 11;  
 Matches 55; Conservative 29; Mismatches 76; Indels 110; Gaps 10;  
 Oy 12 LTFVLMKSLVTHR-LGSGGSDGQFSNVLFSVPLPLPLAGLU--VQOIIFFLGCTALA 68  
 Db 108 LFTFAMSGMFTYALWISGWSNSKY-----FLMGARAVAQIIISYVTLGL 155  
 Oy 69 FLVLMPLVHGNNLFFRS-----LESSWP--FWLTALAA----- 100  
 Db 156 IISNAISLGGVSLMLFTETQEHMMLLSSPLAMMFTSLAETNRSPFDLTGEGSELVS 215  
 Oy 101 -----VILQMAAHVLETHDGHQPLNRRVLAATFLFFPLN 139  
 Db 216 GFNVESAGCFALLFLAEYTNILFNNTLSTMF-----NPGMTNPQ-----LFTIN 262  
 Oy 140 VLVGAMXXXCSPSIAIRHPTPGYVYTRNPLKIEVSQSHPMATFCSLLDAQSLIRPTMA 199  
 Db 263 LMTKTMILT-----LFTMPASTPRFRYDQLMHLMKOYLPULTA 303  
 Oy 200 APQDSLAPGDEBGMQLLOTKDSMAKCAP 229  
 Db 304 -----MYLLNLTSTSMALCGTP 319  
 Db  
 RESULT 15  
 SFUB\_SERMA STANDARD; PRT; 527 AA.  
 ID SFUB\_SERMA  
 AC P21409;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Iron(III)-transport system permease protein sfub.  
 GN SFUB.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxId=615;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90130288; PubMed=2404942;  
 RA Angerer A., Galsner S., Braun V.;  
 RT Nucleotide sequences of the sfua, sfub, and sfuc genes of Serratia

RT marcescens suggest a periplasmic-binding-protein-dependent iron  
 RT transport mechanism";  
 RL J. Bacteriol. 172:572-578 (1990).  
 CC -1- FUNCTION: INVOLVED IN A PERIPLASMIC BINDING-PROTEIN-DEPENDENT  
 CC IRON(III) TRANSPORT SYSTEM, WHICH IS INDEPENDENT OF TONB, EXB8,  
 CC SIDROPHORE, OR ANY OTHER MEMBRANE PROTEIN. HOWEVER, IT IS NOT  
 CC CLEAR HOW FE(III) IS SOLUBILIZED AND TAKEN UP ACROSS THE OUTER  
 CC MEMBRANE.  
 CC -1- SUBUNIT: HOMODIMER (PROBABLE).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT  
 CC SYSTEM PERMEASE FAMILY. CYSTM SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL, M3815; AAA26574.1; -  
 DR PIR: B35108; ORSEUB.  
 DR InterPro: IPR000515; BPD\_transp.  
 DR Pfam: PF00528; BPD\_transp; 2.  
 DR PROSITE: PS00402; BPD\_TRANS\_P-1NN\_MEMBER; FALSE\_NEG.  
 DR Transport; Iron transport; Transmembrane; Inner membrane; Repeat.  
 KM TRANSMEM 27 47  
 FT TRANSMEM 75 95  
 FT TRANSMEM 107 127  
 FT TRANSMEM 139 159  
 FT TRANSMEM 191 211  
 FT TRANSMEM 216 256  
 FT TRANSMEM 284 304  
 FT TRANSMEM 333 353  
 FT TRANSMEM 368 388  
 FT TRANSMEM 396 416  
 FT TRANSMEM 449 469  
 FT TRANSMEM 497 517  
 FT REPEAT 15 15  
 FT REPEAT 66 70  
 FT REPEAT 167 177  
 FT REPEAT 274 280  
 FT REPEAT 319 323  
 FT REPEAT 428 438  
 FT REPEAT 527 AA; 56528 MW; 6B815AD100B0504 CRC64;  
 SQ  
 Query Match 5.0%; Score 84; DB 1; Length 527;  
 Best Local Similarity 23.2%; Pred. No. 19;  
 Matches 80; Conservative 30; Mismatches 133; Indels 102; Gaps 17;  
 Oy 6 LVLSGLTFLVLMRSVTHRLSGSGGSDGQFSNVLFSVPLP-PLAGLVVQOIIFFLGCT 64  
 Db 243 LVLCGLLLLEAISRGARARVAVGSGAR-----SQTRRLSPPLAL--ALLPLAL 294  
 Oy 65 TALA---FLVLMPLVHGNNLFFRSLESWP--FWLTALAVILQMAAHVLETHDGH 119  
 Db 295 TALAAGVPFITLAWLWLGFEVVRNAE-LMPALMOTLSLS----- 334  
 Oy 120 PQLTRRLVLAATFLPLPLVLVGAMXXXCSPSIAIRHPT-----PGY 163  
 Db 335 -----NAGLLITLCAIPMA-----LSRRPARLYRVLBGCNVYTSLSGLIV 377  
 Oy 164 TYRNPFLKIEVSQSHA-MTAFCSLLDAQSLIRPTMAAPQDSLAPGDEBGMQLLOTKDS 222  
 Db 378 VALAVYTTITHSFRPIYOTEITLLAVLMPFRALI-----NLRGIAQAVLELVNARS 433  
 Oy 223 MAKG-----ARPGXRGARWGLATVLTNHTLVQFPKTLILGANGAPCSSL 270  
 Db 434 LKSPAAQALWSTTRRLAPGVAAQA--LVFLAANE-----LRTITLAPNGRTLAT- 485  
 Oy 271 PQSPSITPMAQPPDGHGXAVEVCLHWPXPXGAPPTWMEQ 315



Thu Mar 13 14:11:15 2003

us-09-816-653a-4.rsp

Page 11

Db 486 -GFNALTSEIDYVAAAP--YALIMVAL-----SLPLTWLXSQ 520

Search completed: March 13, 2003, 11:30:26  
Job time : 16.5439 secs

---



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

## OM protein - protein search, using sw model

Run on: March 13, 2003, 11:27:17 ; Search time 18.6069 Seconds

(without alignment)  
1679.148 Million cell updates/sec

Title: US-09-816-653A-4

Perfect score: 1695  
Sequence: 1 CYIALVLSCLTFLVLMRS.....APPTLMESQGFMRKLVG 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613442 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96.5	5.7	593	A11361	probable thiamin A
2	93	5.5	2016	A58195	sodium channel pro
3	91.5	5.4	581	T44942	cytochrome-c oxida
4	90.5	5.3	625	S18420	regulatory protein
5	89.5	5.3	107	D85924	probable cytochrom
6	89.5	5.3	107	C91079	probable cytochrom
7	89.5	5.3	268	S34364	probable cytochrom
8	89	5.3	3164	WMBE6	ul36 protein - hum
9	88.5	5.2	107	A65056	hypothetical prote
10	88.5	5.2	395	AB1061	hypothetical prote
11	87	5.1	621	T49199	growth factor rece
12	86.5	5.1	301	G83556	hypothetical prote
13	86	5.1	528	A10359	iron (iii)-transpor
14	86	5.1	582	A87673	conserved hypothet
15	86	5.1	614	S40085	hypothetical prote
16	85.5	5.0	340	T18704	hypothetical prote
17	85.5	5.0	402	AB0854	hypothetical prote
18	84.5	5.0	166	T01263	probable membrane
19	84.5	5.0	330	T11290	cytochrome-c oxida
20	84.5	5.0	332	B81252	NADH2 dehydrogenas
21	84.5	5.0	377	F95948	probable acyltrans
22	84.5	5.0	636	T35182	probable ABC-type
23	84.5	5.0	1011	T07712	probable ABC-type
24	84	5.0	291	AC1846	sulfate transport
25	84	5.0	321	T11088	NADH2 dehydrogenas
26	84	5.0	477	DB3617	probable amino aci
27	84	5.0	527	Q8SE0B	stib protein - ser
28	84	5.0	585	S75265	probable acid-COA
29	83	4.9	222	S75137	hypothetical prote

30	83	4.9	297	AB2436	permease protein o
31	83	4.9	394	T31891	hypothetical prote
32	83	4.9	487	F72126	ct319 hypothetical
33	83	4.9	596	F75302	ATP-dependent DNA
34	83	4.9	1089	E82987	hypothetical prote
35	83	4.9	1840	CHRTM1	sodium channel pro
36	82.5	4.9	228	T38941	cytochrome-c oxida
37	82.5	4.9	393	B86644	transporter ybf1
38	82.5	4.9	888	JC5399	dual leucine zippe
39	82.5	4.9	3005	T08841	polyprotein - dour
40	82	4.8	335	A87135	probable membrane
41	82	4.8	516	G98198	proline/betaine tr
42	82	4.8	516	AB3088	MFS permease (prol
43	82	4.8	627	A98344	hypothetical prote
44	82	4.8	627	AE2938	conserved hypothet
45	81.5	4.8	315	AG3535	dipeptide transpor

## ALIGNMENTS

## RESULT 1

A71361 probable thiamin ABC transporter, permease protein - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #ext\_change 05-Nov-1999

C:Accession: A71361

R:Fraser, C.M., Norris, S.J., Weinstein, G.M., White, O., Sutton, G.G., Dodson, R., Gwin  
rean, U., Khatkhat, H., Richardson, D., Howell, J.K., Chidambaram, M., Uckerback, T., McDer  
they, L., Weidman, J., Smith, H.O., Venter, J.C.

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: A71361

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-593 <COU>

A:Cross-references: GB:AE001199; GB:AE000520; NID:G3322402; PIDN:AAC65131.1; PID:G332240

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0143

Query Match 5.7% Score 96.5; DB 2; Length 593;  
Best Local Similarity 23.0%; Pred. No. 1.9; Mismatches 86; Indels 89; Gaps 12;  
Matches 61; Conservative 29; Mismatches 86; Indels 89; Gaps 12;

QY	3	ISALVLSCLL----	TFVLMRSVYTHRLSGSGSGDGFQSWNLPSVPLPLPLGLLWQI 58
DB	120	LCSAVLACAVGCAFFFKVKSRLRALFPA-----	LCNIPLAIRPL----- 161
QY	59	IFPLGTTALAPLVMPVLAHGRNLFFRSLSSWPF-----	WTLALA 100
DB	162	-----TWAAFLPLSF-----	GKNGLCRTKLTGTMGVSTPRKFLYSAGCVIAHAWNPFA 212
QY	101	VILQNNAAHVVLETHDGHPOLYNR--RVLYAATFLPLPLVVGAMXXX-----	CSP 151
DB	213	IAIIRAAWNTISADDEADARLIGARAYVFYITL-----	LPALAGALRSSFLVIFLYCFE 267
QY	152	SIIL-----RHPTPGYTYRNFLKIEVSQSH-----	AMTAFCSLL----- 188
DB	268	SLMMVLLLGCTFTTLEVLVRS--IRTOAHFVYSTALSSQTFYALLMIGVSTESQ 324	
QY	189	QAQSLPRTMAAPODSLRPGEDEG 213	
DB	325	QAQSCVARTRPLPTOSIGONIERIG 349	

## RESULT 2

A38195 sodium channel protein phi1, cardiac - human  
N:Alternate names: tetrodotoxin-insensitive, voltage-dependent sodium channel, TTX-I NaCl  
C:Species: Homo sapiens (man)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #ext\_change 20-Aug-1999

C:Accession: A38195  
 R:Geiens, M.E.; George Jr., A.L.; Chen, L.Q.; Chahine, M.; Horn, R.; Barchi, R.L.; Kall  
 Proc. Natl. Acad. Sci. U.S.A. 89, 554-558, 1992  
 A:Title: Primary structure and functional expression of the human cardiac tetrodotoxin-1  
 A:Reference number: A38195; MUID:92115699; PMID:1130946  
 A:Accession: A38195  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-2016 <GEN>  
 A:Cross-references: GB:M77235; NID:G184038; PID:AAA5644.1; PID:G184039  
 A:Experimental source: heart  
 C:Superfamily: sodium channel protein  
 C:Keywords: cardiac muscle; duplication; glycoprotein; heart; ion transport; sodium chan

Query Match 5.5%; Score 93; DB 2; Length 2016;  
 Best Local Similarity 23.8%; Pred. No. 16;  
 Matches 67; Conservative 31; Mismatches 113; Indels 70; Gaps 13;

Db 778 QCGMNIFFD-----SIIVLSIMELGLSRMSNL---SVLRSPRLRVFKLAKSWPTL 825  
 Oy 35 QFSNLFSPVPLPPLAGLVQGITTFGLTALFLVLMPLVLRNLLFRSLSSWPF 94  
 Db 95 -----LTALAVILQNMMAHV-----FETDGHPTLTNR---RVLY 129  
 Oy 826 NTLIKITGNSVGLNLTIVLAIYIFRAVVGMOIFGKNSGLRDSGLPRHMMDF 885  
 Db 130 AATFLPLPLVLMGAKXXCSPSIAIRHPTGYTYRNFLEKIEVSQSHPMATAPCSLLIQ 189  
 Oy 886 HAFLLIF--RLTGEVITETMDQMEVSGS--LCLVFLVLMYIGMLVLMFLALLS 940  
 Db 190 AOSLLPRTMAPDLSRPGEDSGMQLQTKDSMAKGRPAKXGRMGLATLL--HNP 248  
 Oy 941 SFS-----ADNLTAPEDEBENNQL--ALARIQRGLRPAKRTTTFCCGLLRHR 989  
 Db 249 TLQVFRKTLALGANGAP--CSSLPSPSPSITPAMQAPPD 288  
 Oy 990 -----QKPALAAQGLPFCIAIPYSP-----PPE 1015

RESULT 3  
 T44942  
 cytochrome-c oxidase [EC 1.9.3.1] chain I [similarity] - Natronobacterium pharaonis  
 N:Alternate names: baf-type cytochrome-c oxidase chain I  
 C:Species: Natronobacterium pharaonis  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 24-Oct-2000  
 C:Accession: T44942  
 R:Matcar, S.; Engelhard, M.  
 Eur. J. Biochem. 250, 332-341, 1997  
 A:Title: Cytochrome baf from Natronobacterium pharaonis: An archaeal four-subunit cytoch  
 A:Reference number: 222876; MUID:98088958; PMID:9428662  
 A:Accession: T44942  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-581 <MAT>  
 A:Cross-references: EMBL:Y10500; PIDN:CAA71525.1  
 A:Experimental source: strain SP1/28  
 C:Genetic:

A:Gene: cbaA  
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology  
 C:Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated comp  
 F/228-332/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted  
 F/379/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 5.4%; Score 91.5; DB 2; Length 581;  
 Best Local Similarity 26.6%; Pred. No. 5;  
 Matches 63; Conservative 32; Mismatches 89; Indels 53; Gaps 15;

Db 6 LVSCLTFLVLRSLVTRHLSGGSGDQPSWNLFSVPLPLPAGLVQGITTFGLTT 65  
 Oy 300 LLLPSLLTATVVAASMEHGARGGGS--GYFGM--LRALPMRDPVFTGMALAGLMF---A 352  
 Db 66 ALAPLVLMPLVLRNLLFRSLSSW---PWLTLALAVILQNMMAHVFLTHDHP 121

Db 353 AAAFSGM--VMGMNINYL--VHNTMWVGHFLTGTAVALLFMVAVSWFL-----PQ 402  
 Oy 122 LTNRR-----VLNATFLIF--PLNVLMGAKXXCSPSIAIRHPTGYTYRNFLEKIEVSO 175  
 Db 403 ITQKLMGKSVLAQVLMFVGMETMSNMAHRSGLAGMPRTAEF--QYRN--EEEMA 458  
 Oy 176 SH-----PAMTAFCSLLLOQSLL-----PRTMAPO--DSLRGEED 211  
 Db 459 GSIGELNAQVAVGILLFVSTILLFVLVVMVTVLDKAEPEGLRANEADTL--SGPD 514

RESULT 4  
 S18420  
 regulatory protein nifA - Azospirillum brasiliense  
 C:Species: Azospirillum brasiliense  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 26-Aug-1999  
 C:Accession: S18420  
 R:Liang, Y.Y.; Kaminski, P.A.; Elmerich, C.  
 Mol. Microbiol. 5, 2735-2744, 1991  
 A:Title: Identification of a nifA-like regulatory gene of Azospirillum brasiliense Sp7 exi  
 A:Reference number: S18420; MUID:92140038; PMID:1779763  
 A:Accession: S18420  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-625 <LTA>  
 A:Cross-references: EMBL:X60714; NID:G38677; PIDN:CAA43126.1; PID:G38678  
 C:Genetic:

A:Gene: nifA  
 C:Superfamily: nif-specific regulatory protein; RNA polymerase sigma factor interaction  
 C:Keywords: DNA binding; P-loop; transcription regulation  
 F/205-426/Domain: RNA polymerase sigma factor interaction domain homology <SFI>  
 F/233-240/Region: nucleotide-binding motif A (P-loop) #status atypical  
 F/300-304/Region: nucleotide-binding motif B

Query Match 5.3%; Score 90.5; DB 2; Length 625;  
 Best Local Similarity 27.7%; Pred. No. 6.7;  
 Matches 48; Conservative 9; Mismatches 55; Indels 6; Gaps 9;

Db 184 CSLLLOQSLL--PRTMAPDLSRPGEDSGMQLQTKDSMAKGA-----RPG----- 230  
 Db 447 CSLLNCGSSVLPQYRTLAGSVGLAP-----SMQPGALINRVPGRPGCPAA 492  
 Oy 231 -----AKGRARWGLAYTLNHPITQVFRKTLALGANG-----AQP 266  
 Db 493 ANAPKTPAMPAPVPEBAGAAARGRP---ARRVPRPLAGLRARRA--CGSGPDPACP 546  
 Oy 267 CSSLPSPSPSITPAMQAPRPHKXGVVCLHMPXXGSAAPTWMESQGF 319  
 Db 547 CPERARLPQAPRPPSPAARPP---AAEVPLD--EPESGLRDRLNLMERTGW 595

RESULT 5  
 D85924  
 probable cytochrome oxidase subunit ygbE [imported] - Escherichia coli (strain O157:H7, E  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: D85924  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
 111er, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potluri, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: D85924  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-107 <STO>  
 A:Cross-references: GB:AB005174; NID:G12517203; PIDN:AA657856.1; GSPDB:GN00145; UWGP:Z405  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetic:

A:Gene: ygbE

Query Match 5.3%; Score 89.5; DB 2; Length 107;  
 Best Local Similarity 24.6%; Pred. No. 1.1;

Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;

QY 33 DQGFNNLFSVPLPLPPLAGLVQOIFFLGTALAVLMPVL-HGRNLLFFRSLESSW 91  
 DB 18 DEETWS-----LPGAVV-----GFISWLFALAMPMLIGSNTLFF--FIYTW 58

QY 92 PFWLTLALAVIIONMAAHVFLETHDGPOLNRRVLVYATFLPLPLVGVGM 145  
 DB 59 PFWLALMPVAVVGVIALH-----SLMDGKLKRSIVFTLVTGIVFGAL 101

RESULT 6  
 C91079  
 probable cytochrome oxidase subunit [imported] - Escherichia coli (strain O157:H7, subsp  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C/Accession: C91079  
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 gasawara, N.; Yasunaga, T.; Kuahara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc  
 A/Reference number: A59623; MUID:21156231; PMID:11258796  
 A/Accession: C91079  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-107 <HA>  
 A/Cross-references: GB:BA000007; PIDN:BA037026.1; PID:g13363074; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain RIMD 0509952  
 C/Genetics:  
 A/Gene: EC03603

Query Match 5.3%; Score 89.5; DB 2; Length 107;  
 Best Local Similarity 24.6%; Pred. No. 1.1; Mismatches 34; Indels 31; Gaps 5;  
 Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;

QY 33 DQGFNNLFSVPLPLPPLAGLVQOIFFLGTALAVLMPVL-HGRNLLFFRSLESSW 91  
 DB 18 DEETWS-----LPGAVV-----GFISWLFALAMPMLIGSNTLFF--FIYTW 58

QY 92 PFWLTLALAVIIONMAAHVFLETHDGPOLNRRVLVYATFLPLPLVGVGM 145  
 DB 59 PFWLALMPVAVVGVIALH-----SLMDGKLKRSIVFTLVTGIVFGAL 101

RESULT 7  
 S34364  
 hypothetical protein X - Salmonella typhimurium (fragment)  
 C/Species: Salmonella typhimurium  
 C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 08-Oct-1999  
 C/Accession: S34364  
 R;Persson, B.C.; Bjork, G.R.  
 submitted to the EMBL Data Library, June 1993  
 A/Description: Isolation of a gene (miaB) involved in hydroxylation of ms216a in CRNA c  
 A/Reference number: S34359  
 A/Accession: S34364  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-268 <PER>  
 A/Cross-references: EMBL:X73368; NID:g1312705; PIDN:CA051785.1; PID:g1312711

Query Match 5.3%; Score 89.5; DB 2; Length 268;  
 Best Local Similarity 26.2%; Pred. No. 3.1; Mismatches 45; Indels 39; Gaps 6;  
 Matches 37; Conservative 20; Mismatches 45; Indels 39; Gaps 6;

QY 5 ALVLSCLTFLVLRSLVTHRIGSGSGDQGFNNLFSVPLPLPPLAGLVQOIFFLKST 64  
 DB 27 ALPISVLVFIIVISLSLTHGHPGIG-----FTLF-----GLLI-ALIPFMAV 70

QY 65 TALAFLVLMPLVHGNLLFFRSLESSWFWLTLALAVIIONMAAHVFLETHDGPOLN 124  
 DB 71 KQLQVQAMTSLINGVHFGFQCSMRRAW--WYMFALPVL--WVA----- 110

QY 125 RRVLVAAATFLPLPLVGVGM 145

DB 111 ---LYIVLYIISLVTAIVGSL 128

RESULT 8  
 WMBE6  
 UL36 protein - human herpesvirus 1 (strain 17)  
 C/Species: human herpesvirus 1  
 C/Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jun-2000  
 C/Accession: 130085  
 R;McGeoch, D.J.; Dairymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perry  
 J.; Gen. Virol. 69, 1531-1574, 1988  
 A/Title: The complete DNA sequence of the long unique region in the genome of herpes sim  
 A/Reference number: A30083; MUID:88274327; PMID:2839594  
 A/Accession: 130085  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-316 <CMCG>  
 A/Cross-references: GB:X14112; NID:g1944536; PIDN:CA032311.1; PID:g959536; GB:D00317  
 C/Genetics:  
 A/Gene: UL36  
 C/Superfamily: varicella-zoster virus gene 22 protein

Query Match 5.3%; Score 89; DB 1; Length 3164;  
 Best Local Similarity 36.9%; Pred. No. 57; Mismatches 17; Indels 20; Gaps 3;  
 Matches 24; Conservative 4; Mismatches 17; Indels 20; Gaps 3;

QY 247 NPTLOVFRKTKLLGANGAQCSGLPSPSITPMAQAPDPHXGVVCLHWPXXGSA 306  
 DB 2824 DPTAVP-----LCRNPEEPTSSPACPSPPPAVQVAPVAP-----PTSG-P 2863

QY 307 PPTWL 311  
 DB 2864 PPTVL 2868

RESULT 9  
 A65056  
 hypothetical protein in surf-cyc intergenic region - Escherichia coli (strain K-12)  
 C/Species: Escherichia coli  
 C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C/Accession: A65056  
 R;Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, M.T.; Burland, V.; Riley, M.; Coi  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A/Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; MUID:97426617; PMID:9278503  
 A/Accession: A65056  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-107 <BLAT>  
 A/Cross-references: GB:A6000358; GB:U00096; NID:g2367156; PIDN:AA075791.1; PID:g1789106;  
 A/Experimental source: strain K-12, substrain MG1655  
 C/Genetics:  
 A/Gene: Y9DE

Query Match 5.2%; Score 88.5; DB 2; Length 107;  
 Best Local Similarity 24.6%; Pred. No. 1.3; Mismatches 34; Indels 31; Gaps 5;  
 Matches 28; Conservative 21; Mismatches 34; Indels 31; Gaps 5;

QY 33 DQGFNNLFSVPLPLPPLAGLVQOIFFLGTALAVLMPVL-HGRNLLFFRSLESSW 91  
 DB 18 DEETWS-----LPGAVV-----GFISWLFALAMPMLIGSNTLFF--FIYTW 58

QY 92 PFWLTLALAVIIONMAAHVFLETHDGPOLNRRVLVYATFLPLPLVGVGM 145  
 DB 59 PFWLALMPVAVVGVIALH-----SLMDGKLKRSIVFTLVTGIVFGAL 101

RESULT 10  
 AB1061  
 probable inner membrane protein STY403 [imported] - Salmonella enterica subsp. enteric.  
 C/Species: Salmonella enterica subsp. enterica serovar Typhi

**RESULT 12**

hgb3556  
C/Spectrum: Pseudomonas aeruginosa (strain PA01)  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/accession: G83556  
R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A/reference number: AB2950; MID:2043737; PMID:10984043  
A/accession: G83556  
A/status: preliminary  
A/molecule type: DNA  
A/residues: 1-301 <STO>  
A/cross-references: GB:AE004506; GB:AE004091; NID:99946584; PIDN:AA04091.1; GSPDB:GN001  
A/experimental source: strain PA01  
C/genetics:  
A/gene: PA0702

Query Match 5.1%; Score 86.5; DB 2; Length 301;  
Best Local Similarity 21.9%; Pred. No. 6.3;  
Matches 61; Conservative 27; Mismatches 105; Indels 85; Gaps 10;

OY 58 IFFPQTALATLV-----LMPVHGCR-----NIL----- 82  
| | | | | | | | | | | | | | | | | |  
Db 36 VALFIGAGLGSLMERLPYQPRWRPHADRLRVYLASVNESLNGLIALPLGVLG 95  
| | | | | | | | | | | | | | | | | |

OY 83 PRSLSESWPFRLTALAVALIOMNAHWPLETDGHQPLTNRRVLYATFLLFPNLV 142  
| | | | | | | | | | | | | | | | | |  
Db 96 FWPVPQMPLMLQLLAIVLADIGTLVHVASHT-----SSALLMRHH--- 138  
| | | | | | | | | | | | | | | | | |

OY 143 GAMXXXCSPSIAIRHPPTPGYYTYRNFLKIEVSOSHPMATFSCLLDAQSLPRTMAAPQ 202  
| | | | | | | | | | | | | | | | | |  
Db 139 -----AVHSVORLYGFNGLMKHPLHLGLEALGSTPLPLLGG--VPQTVAALL 184  
| | | | | | | | | | | | | | | | | |

OY 203 DLRGEDEGMOLLQTKDSWAKGARPGARXGRARWCILATTLH-----NPTLOVF 253  
| | | | | | | | | | | | | | | | | |  
Db 185 AFA-----IQTILLQSHNV--DMRGIRHYFAAPAFLRHRIHYAGADVAFALPFS 236  
| | | | | | | | | | | | | | | | | |

OY 254 RKTALIGANGAQPCSSILPGSPPTITPAMOPAGPPDHGX 291  
| | | | | | | | | | | | | | | | | |  
Db 237 VMDRLGTALHRPDYRDLDSTMGI--GDQPDPYRDYAG 272  
| | | | | | | | | | | | | | | | | |

**RESULT 13**

iron(III)-transport system permease sfub [imported] - Yersinia pestis (strain CO92)  
AIO359  
C/Spectrum: Yersinia pestis  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C/accession: AI0359  
Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.; den, Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.; de la Maza, L.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, T.  
Nature 413, 523-527, 2001

A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague  
A/reference number: AB0001; MUID:21470413; PMID:11586360  
A/accession: AI0359  
A/status: preliminary  
A/molecule type: DNA  
A/residues: 1-528 <CUR>  
A/cross-references: CB:AL590842; PIDN:CAC92204.1; PID:915980916; GSPDB:GN00175  
C/genetics:  
A/gene: sfub  
C/superfamily: sfub protein

Query Match 5.1%; Score 86; DB 2; Length 528;  
Best Local Similarity 22.3%; Pred. No. 13;  
Matches 78; Conservative 31; Mismatches 129; Indels 112; Gaps 15;

Db 244 LALCCLGLLVEATSGRYRVARVSGTPT-----ROTYSMGTS 283  
Qy 66 ALAFVLAMVLMGRNLLFFRSLESSMP-----WLT-----LAL 99  
Db 284 LTLCLLPL-----ITTLGCVPTTLMRWLSGGCIDIMLNPPELLPALKQTLGLML 336  
Qy 100 --AVILQNNAAHMFLETHDGHPOLTNRVLYAATFLPLVNLVGMXXCSPSIAIRH 157  
Db 337 SGAVITTLCAIPMAWLSVR--YPCRLHR-----AMEGCYYTSSIPGIVVLAIVTITIRI 390  
Qy 158 PPGGYTYTNFLKIEVSQSHPMPTFCSLILOAQSLLPRTMARPOSLRGEDEGMOLL 217  
Db 391 ARPLY-----QTEFTVLLVYLMFTPRAT-----SLRAGIAQARVEL 429  
Qy 218 QTKDSMAKG-----ARPGAKRGARMGCLAYTLHNPTLOVFRKTALLGANGAO 265  
Db 430 NVARSIGRPTQAMLTTLRLAFAAGAA---LVFLAISNE---LTATLLADNGTR 482  
Qy 266 POSLPGSPSITPAMOPAGPBDHXGVYVCLHWPXKXGAPPTWLMESQ 315  
Db 483 TLAT--GFWALTSBIDVYAAB--YAFIMVNL-----SLPTWLLYSQ 521

## RESULT 14

A87673  
conserved hypothetical protein CC3419 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: A87673  
R:Netman, W.C.; Feldbljum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eiken, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaev, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: A87673  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-582 <STO>  
A:Cross-references: GB:AE005673; NID:G13425133; PIDN:AAK25381.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC3419

Query Match 5.1%; Score 86; DB 2; Length 582;  
Best Local Similarity 24.3%; Pred. No. 15;  
Matches 74; Conservative 30; Mismatches 113; Indels 88; Gaps 14;

Qy 62 LGTTALAFVLMVLMGRNLLFFRSLESSMPWLTALAVILQNNAAHMFVFL--ETHDG 118  
Db 220 VGLTLGMLPLMNLGGGFLATLPFTGLAPLMGTAKATALLSAAAALLILINAAVQDG 279  
Qy 119 H--POLTNRRLVYAATFLPLVNLVY-----AMXXXCSP 151  
Db 280 EBPPLIIPCLAIARIGLILIPVILAGYALMLRIDOYGLTBRVAVGVLVVAIGFTAGY 339  
Qy 152 SIAIRHTPGYTYTNFLKIEVSQSHPMPTAFCSLL-----QAQSLL 194  
Db 340 ALAAVAKGP-----WMK-PLERTNPIMAAACVLLILALFTPIANPARLSVASQVRUE 391  
Qy 195 PRMAAPQ-----DSLPGEBDEGMOLLQTKDS--MAGARPGAKRGARGLAYTL 244  
Db 392 SGVAADKDFDPLRPDARVGR--EALDRLKTHPMARAIARRDAA-----ASTE 440  
Qy 245 LHNPTLOV--FRKTALLGANGAOPCS-----SLPGSPSITPAMQ-PA--GPPDHGX 292  
Db 441 KQYPAGEIRPDFAMAVYPAKALPQSPFAQDWQSPSGSNCTAIAMQCPLVADVADGK 500  
Qy 293 VEYCL 297  
Db 501 DEYLL 505

RESULT 15

S40085  
hypothetical protein 4 - Lactococcus lactis

C:Species: Lactococcus lactis  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 15-Oct-1999  
C:Accession: S40085  
R:Batton, T.D.; Shearman, C.; Gasson, M.  
submitted to the EMBL Data Library, December 1993  
A:Description: Cloning and sequence analysis of the DNAK gene region of Lactococcus lact  
A:Reference number: S40082  
A:Accession: S40085  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-614 <BAT>  
A:Cross-references: EMBL:X76642; NID:G435489; PIDN:CAA54090.1; PID:G435493

Query Match 5.1%; Score 86; DB 2; Length 614;  
Best Local Similarity 19.9%; Pred. No. 16;  
Matches 49; Conservative 35; Mismatches 74; Indels 88; Gaps 11;

Qy 5 ALVLSCLLFLVLMVLMGRNLLFFRSLESGSGDQGFNSLFSV-----PLPLPLAGLLVQOI 58  
Db 332 SLTFSQSLVLMQPL-----FPMKLSITTYQDPARSTLFGFL---- 374  
Qy 59 IFFLTALAFVLMVLMGRNLLFFRSLESSMPWLTALAVILQNNAAHMFLETHDG 118  
Db 375 -----SALSLVILPIL-----LDKISGKTSYVTLIGLIVFSL-----G 410  
Qy 119 HPOLTR-----RVLYAA-----TFLPLVNLVGMXXCSPSIAIRHTPGY 163  
Db 411 FAEFRRIKSGQPLFASQSLNKTPTNYMENDSTAIGBYL---FOVIGSHNOPYK 466  
Qy 164 TYNPLKIE--VSQSHPMPTAFCSLLLOAQSLLPRTMARPOSLRGEDEGMOLLQTYD 221  
Db 467 TIQGFKDKNVYGVGMNQAMT-----YLSRGKLPBGCLA-----KSIQISD 506  
Qy 222 SMAKGA 227  
Db 507 YSKXGS 512

Search completed: March 13, 2003, 11:31:03  
Job time : 23.6069 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

## CM protein - protein search, using sw model

Run on: March 13, 2003, 11:30:32 ; Search time 13.645 Seconds  
(without alignments)  
1097.829 Million cell updates/sec

Title: US-09-816-653a-4  
Perfect score: 1695  
Sequence: 1 CYISALVLSCLTFLVLMRS.....APPTLMESQGFMRKLVG 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues  
Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US60\_PUB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1679	99.1	325	US-09-816-653a-4	Sequence 4, App1
2	1110.5	65.5	667	US-09-759-056-2	Sequence 2, App1
3	1110.5	65.5	667	US-09-901-812-2	Sequence 2, App1
4	1110.5	65.5	667	US-10-227-884-80	Sequence 80, App1
5	1110.5	65.5	667	US-10-230-163-80	Sequence 80, App1
6	1110.5	65.5	667	US-10-218-631-80	Sequence 80, App1
7	1110.5	65.5	667	US-10-230-338-80	Sequence 80, App1
8	1106.5	65.3	658	US-09-759-056-5	Sequence 5, App1
9	1106.5	65.3	658	US-09-901-812-5	Sequence 5, App1
10	832.5	49.1	670	US-09-816-653a-7	Sequence 7, App1
11	485	28.6	178	US-09-925-299-796	Sequence 796, App
12	485	28.6	178	US-09-925-299-796	Sequence 796, App
13	97.5	5.8	403	US-09-826-508-30	Sequence 30, App1
14	97.5	5.8	403	US-09-895-686-5	Sequence 5, App1
15	97.5	5.8	427	US-09-826-508-32	Sequence 32, App1
16	95.5	5.6	526	US-09-738-626-5712	Sequence 5712, App
17	93	5.4	2016	US-09-840-125-4	Sequence 4, App1
18	92	5.4	664	US-09-767-870-9	Sequence 9, App1
19	92	5.4	2100	US-09-995-542-6	Sequence 6, App1

20	92	5.4	2144	10	US-09-858-194-2	Sequence 2, App1
21	92	5.4	2146	10	US-09-995-542-5	Sequence 5, App1
22	91.5	5.4	572	10	US-09-919-781-2	Sequence 2, App1
23	89	5.3	249	9	US-09-738-626-6183	Sequence 6183, App
24	87	5.1	621	9	US-10-242-332-4	Sequence 4, App1
25	87	5.1	2121	10	US-09-995-542-3	Sequence 3, App1
26	87	5.1	2121	10	US-09-995-542-2	Sequence 2, App1
27	84.5	5.0	299	9	US-09-510-332-7	Sequence 7, App1
28	84.5	5.0	1508	9	US-09-393-634-41	Sequence 41, App1
29	84.5	5.0	1508	9	US-10-024-623-35	Sequence 35, App1
30	84.5	5.0	1508	12	US-10-002-769-15	Sequence 15, App1
31	84	5.0	688	9	US-10-174-529-524	Sequence 524, App
32	84	5.0	688	9	US-10-176-758-524	Sequence 524, App
33	84	5.0	688	9	US-10-175-737-524	Sequence 524, App
34	84	5.0	688	9	US-10-173-706-524	Sequence 524, App
35	84	5.0	688	9	US-10-175-728-524	Sequence 524, App
36	84	5.0	688	9	US-10-175-752-524	Sequence 524, App
37	84	5.0	688	9	US-10-176-482-524	Sequence 524, App
38	84	5.0	688	9	US-10-176-913-524	Sequence 524, App
39	84	5.0	688	9	US-10-180-532-524	Sequence 524, App
40	84	5.0	688	9	US-10-180-557-524	Sequence 524, App
41	84	5.0	688	9	US-10-173-700-524	Sequence 524, App
42	84	5.0	688	9	US-10-174-570-524	Sequence 524, App
43	84	5.0	688	9	US-10-174-579-524	Sequence 524, App
44	84	5.0	688	9	US-10-174-579-524	Sequence 524, App
45	84	5.0	688	9	US-10-174-582-524	Sequence 524, App

## ALIGNMENTS

```
RESULT 1
US-09-816-653a-4
Sequence 4, Application US/0981653a
Publication No. US20030021788A1
GENERAL INFORMATION:
APPLICANT: Kastlella, Luca
TITLE OF INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: 10716/57
CURRENT APPLICATION NUMBER: US/09/816,653a
PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 60/191,532
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 325
TYPE: PRT
ORGANISM: Homo sapiens hstra6 polypeptide fragment, carboxy terminus
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (146)..(148)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (232)..(232)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (290)..(290)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (292)..(292)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (302)..(302)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
NAME/KEY: MISC FEATURE
LOCATION: (303)..(303)
OTHER INFORMATION: Xaa represents any amino acid, other, or unknown
US-09-816-653a-4
Query Match 99.1%; Score 1679; DB 9; Length 325;
Best Local Similarity 100.0%; Pred. No. 7; 6e-143;
```

Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CYISALVLSCLTFLVLMRSVLTNRALSGSGSGGQFQSNLFSVPLPLPPLAGLLVQOITIF 60  
DB 1 CYISALVLSCLTFLVLMRSVLTNRALSGSGSGGQFQSNLFSVPLPLPPLAGLLVQOITIF 60

QY 61 FLGTTALAFVLVLMFVLMHGRNLLFPRSLSSMPFWLTALAVILQNMMAHWVLETHDGH 120  
DB 61 FLGTTALAFVLVLMFVLMHGRNLLFPRSLSSMPFWLTALAVILQNMMAHWVLETHDGH 120

QY 121 QLTNRVLYAATFLPLPLVLMVGMAMXXCSPSIAIHNPFGYTYTRNFKIEVSQSHPM 180  
DB 121 QLTNRVLYAATFLPLPLVLMVGMAMXXCSPSIAIHNPFGYTYTRNFKIEVSQSHPM 180

QY 181 TAFCSLLLOAQOSLIPRTMAAPOSLSRGEDESGMQLQTKSMAGARPGARARWGL 240  
DB 181 TAFCSLLLOAQOSLIPRTMAAPOSLSRGEDESGMQLQTKSMAGARPGARARWGL 240

QY 241 AVTLNHPDLOVFRKTAALLGANGAQCCSLPSGSPSITPAPQAPGPPHGXVVCVCLHME 300  
DB 241 AVTLNHPDLOVFRKTAALLGANGAQCCSLPSGSPSITPAPQAPGPPHGXVVCVCLHME 300

QY 301 PXXGSAAPPVLMESQGFMRKCLVG 325  
DB 301 PXXGSAAPPVLMESQGFMRKCLVG 325

## RESULT 2

US-09-759-056-2  
Sequence 2, Application US/09759056  
Patent No. US20020156252A1  
GENERAL INFORMATION:  
APPLICANT: Pennica, Diane  
APPLICANT: Smith, Victoria  
TITLE OF INVENTION: No. US20020156252A1 STRA6 Polypeptides  
FILE REFERENCE: GENENT.289782  
CURRENT APPLICATION NUMBER: US/09/759,056  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/197089  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/175849  
PRIOR FILING DATE: 2000-01-13  
PRIOR APPLICATION NUMBER: 60/228914  
PRIOR FILING DATE: 2000-08-29  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 667  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-09-759-056-2

Query Match 65.5%; Score 1110.5; DB 9; Length 667;  
Best Local Similarity 76.6%; Pred. No. 1.6e-91;

Matches 222; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGD-----GQFSWNLFS 42  
DB 365 CYISALVLSCLTFLVLMRSVLTNRALHGRALDLSPLHRSPPHRSQALFCMSMS 424

QY 43 VPLPLPPLAGLLVQOITIFLGTALAFVLMFVLMHGRNLLFPRSLSSMPFWLTALAVI 102  
DB 425 AVGTATICGLVQOITIFLGTALAFVLMFVLMHGRNLLFPRSLSSMPFWLTALAVI 484

QY 103 LQNMMAHWVLETHDGHQLTNRVLYAATFLPLPLVLMVGMAMXXC----- 149  
DB 485 LQNMMAHWVLETHDGHQLTNRVLYAATFLPLPLVLMVGMAMXXC----- 149

QY 150 -SPSIAIRHP-----TPGYTYTRNFKIEVSQSHPMATFCSLLLOAQOSLIPRTMAAPD 203  
DB 545 GQMDLSLPPRAATLDPGYTYTRNFKIEVSQSHPMATFCSLLLOAQOSLIPRTMAAPD 604

QY 204 SLRGEDESGMQLQTKDSMAGARPGARARWGLAVTLNHPDLOVFRKTAALLGANG 263  
DB 605 SLRGEDESGMQLQTKDSMAGARPGARARWGLAVTLNHPDLOVFRKTAALLGANG 664

QY 264 AQP 266  
DB 665 AQP 667

## RESULT 3

US-09-901-812-2  
Sequence 2, Application US/09901812  
Patent No. US20020173461A1  
GENERAL INFORMATION:  
APPLICANT: Pennica, Diane  
APPLICANT: Polakis, Paul  
APPLICANT: Szeto, Wayne  
TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO  
FILE REFERENCE: ENHANCE EFFICACY OF IMMUNOTHERAPY  
CURRENT APPLICATION NUMBER: US/09/901,812  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/228,914  
PRIOR FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 09/759,056  
PRIOR FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/175,849  
PRIOR FILING DATE: 2000-01-13  
PRIOR APPLICATION NUMBER: 60/197,089  
PRIOR FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 667  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-901-812-2

Query Match 65.5%; Score 1110.5; DB 9; Length 667;  
Best Local Similarity 76.6%; Pred. No. 1.6e-91;

Matches 222; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLTNR-----LGSQSGD-----GQFSWNLFS 42  
DB 365 CYISALVLSCLTFLVLMRSVLTNRALHGRALDLSPLHRSPPHRSQALFCMSMS 424

QY 43 VPLPLPPLAGLLVQOITIFLGTALAFVLMFVLMHGRNLLFPRSLSSMPFWLTALAVI 102  
DB 425 AVGTATICGLVQOITIFLGTALAFVLMFVLMHGRNLLFPRSLSSMPFWLTALAVI 484

QY 103 LQNMMAHWVLETHDGHQLTNRVLYAATFLPLPLVLMVGMAMXXC----- 149  
DB 485 LQNMMAHWVLETHDGHQLTNRVLYAATFLPLPLVLMVGMAMXXC----- 149

QY 150 -SPSIAIRHP-----TPGYTYTRNFKIEVSQSHPMATFCSLLLOAQOSLIPRTMAAPD 203  
DB 545 GQMDLSLPPRAATLDPGYTYTRNFKIEVSQSHPMATFCSLLLOAQOSLIPRTMAAPD 604

QY 204 SLRGEDESGMQLQTKDSMAGARPGARARWGLAVTLNHPDLOVFRKTAALLGANG 263  
DB 605 SLRGEDESGMQLQTKDSMAGARPGARARWGLAVTLNHPDLOVFRKTAALLGANG 664

QY 264 AQP 266  
DB 665 AQP 667

## RESULT 4

US-10-227-884-80  
Sequence 80, Application US/10227884  
Publication No. US20030027988A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.

APPLICANT: Desnoyers, Luc  
APPLICANT: Gerritsen, Mary  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gunney, Austin L.  
APPLICANT: Smith, Victoria  
APPLICANT: Stephan, Jean-Philippe F.  
APPLICANT: Watanabe, Colin L.  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3530P1C79  
CURRENT APPLICATION NUMBER: US/10/227,884  
CURRENT FILING DATE: 2002-08-26  
PRIOR APPLICATION NUMBER: 10/119,480  
PRIOR FILING DATE: 2002-04-09  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/062287  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063549  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/064103  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/069873  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089905  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090691  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/095302  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/095318  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/095916  
PRIOR FILING DATE: 1998-08-10  
PRIOR APPLICATION NUMBER: 60/096146  
PRIOR FILING DATE: 1998-08-11

PRIOR APPLICATION NUMBER: 60/096791  
PRIOR FILING DATE: 1998-08-17  
PRIOR APPLICATION NUMBER: 60/097986  
PRIOR FILING DATE: 1998-08-26  
PRIOR APPLICATION NUMBER: 60/098544  
PRIOR FILING DATE: 1998-08-31  
PRIOR APPLICATION NUMBER: 60/099596  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099803  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099811  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099812  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099816  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100038  
PRIOR FILING DATE: 1998-09-11  
PRIOR APPLICATION NUMBER: 60/100385  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100627  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100848  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/100919  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/101477  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101738  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101741  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101786  
PRIOR FILING DATE: 1998-09-25  
PRIOR APPLICATION NUMBER: 60/101916  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101922  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/106178  
PRIOR FILING DATE: 1998-10-28  
PRIOR APPLICATION NUMBER: 60/106248  
PRIOR FILING DATE: 1998-10-29  
PRIOR APPLICATION NUMBER: 60/106464  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: 60/106905  
PRIOR FILING DATE: 1998-11-03  
PRIOR APPLICATION NUMBER: 60/108787  
PRIOR FILING DATE: 1998-11-17  
PRIOR APPLICATION NUMBER: 60/108801  
PRIOR FILING DATE: 1998-11-17  
PRIOR APPLICATION NUMBER: 60/108849  
PRIOR FILING DATE: 1998-11-18  
PRIOR APPLICATION NUMBER: 60/112422  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113296  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/113605  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113621  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/115558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119549  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/123618

PRIOR FILING DATE: 1999-03-10  
PRIOR APPLICATION NUMBER: 60/125259  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 60/125775  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/126773  
PRIOR FILING DATE: 1999-03-29  
PRIOR APPLICATION NUMBER: 60/127887  
PRIOR FILING DATE: 1999-04-05  
PRIOR APPLICATION NUMBER: 60/130232  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/131022  
PRIOR FILING DATE: 1999-04-26  
PRIOR APPLICATION NUMBER: 60/131270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131445  
PRIOR FILING DATE: 1999-04-28  
PRIOR APPLICATION NUMBER: 60/134287  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/140650  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/140723  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/141037  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/144758  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/145698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 60/146222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: 60/146963  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/149320  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/149638  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/151733  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: 60/164418  
PRIOR FILING DATE: 1999-11-09  
PRIOR APPLICATION NUMBER: 60/166361  
PRIOR FILING DATE: 1999-11-16  
PRIOR APPLICATION NUMBER: 60/169445  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: 60/169495  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: 60/169835

Query Match 65.5% Score 110.5; DB 9; Length 667;  
Best Local Similarity 76.6% Pred. No. 1.6e-91;  
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFLVLMKSLVTHR-----LGSQSGSD-----GQFSNLF 42  
DB 365 CYSALVLSCLTFLVLMKSLVTHR-----LGSQSGSD-----GQFSNLF 424  
QY 43 VPLPLPLAGLLVQOIIFFELGTTALAFVLMPVLHGRNLLFFSLSBSPFMTALAVI 102  
DB 425 AVGTAFICGLLVQOIIFFELGTTALAFVLMPVLHGRNLLFFSLSBSPFMTALAVI 484  
QY 103 LQNMAMHWFLETHGHQOLNRRVLYAATLFLPLVAVLYGAKXXXC----- 149  
DB 485 LQNMAMHWFLETHGHQOLNRRVLYAATLFLPLVAVLYGAKXXXC----- 149  
QY 150 -SPSIAIRHP-----TPGVYTYRNFLKIEVSQSHPMATAFCSLLLOASLLPTMAAPD 203  
DB 545 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMATAFCSLLLOASLLPTMAAPD 604  
QY 204 SLRPGEDSGMQLLOTKDSMAKARGASRGARWGIAVYTLNHPVLQVFRKTALLGANG 263

DB 605 SLRPGEDSGMQLLOTKDSMAKARGASRGARWGIAVYTLNHPVLQVFRKTALLGANG 664  
QY 264 AQP 266  
DB 665 AQP 667  
RESULT 5  
US-10-230-163-80  
Sequence 80, Application US/10230163  
Publication No. US20030036635A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Desnoyers, Luc  
APPLICANT: Gerritsen, Mary  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Guirney, Austin L.  
APPLICANT: Smith, Victoria  
APPLICANT: Stephan, Jean-Philippe F.  
APPLICANT: Watanabe, Colin L.  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
ACIDS ENCODING THE SAME  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
ACIDS ENCODING THE SAME  
CURRENT APPLICATION NUMBER: US/10/230,163  
CURRENT FILING DATE: 2002-08-28  
PRIOR APPLICATION NUMBER: 10/119,480  
PRIOR FILING DATE: 2002-04-09  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/062287  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063549  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/064103  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/069873  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089905  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090691

PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/095302  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/095318  
PRIOR FILING DATE: 1998-08-04  
PRIOR APPLICATION NUMBER: 60/095916  
PRIOR FILING DATE: 1998-08-10  
PRIOR APPLICATION NUMBER: 60/096146  
PRIOR FILING DATE: 1998-08-11  
PRIOR APPLICATION NUMBER: 60/096791  
PRIOR FILING DATE: 1998-08-17  
PRIOR APPLICATION NUMBER: 60/097986  
PRIOR FILING DATE: 1998-08-26  
PRIOR APPLICATION NUMBER: 60/098544  
PRIOR FILING DATE: 1998-08-31  
PRIOR APPLICATION NUMBER: 60/099596  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099598  
PRIOR FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: 60/099803  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099811  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099812  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/099816  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100038  
PRIOR FILING DATE: 1998-09-11  
PRIOR APPLICATION NUMBER: 60/100385  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/100627  
PRIOR FILING DATE: 1998-09-16  
PRIOR APPLICATION NUMBER: 60/100848  
PRIOR FILING DATE: 1998-09-18  
PRIOR APPLICATION NUMBER: 60/100919  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: 60/101477  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/101738  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101741  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101786  
PRIOR FILING DATE: 1998-09-25  
PRIOR APPLICATION NUMBER: 60/101916  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/101922  
PRIOR FILING DATE: 1998-09-24  
PRIOR APPLICATION NUMBER: 60/106178  
PRIOR FILING DATE: 1998-10-28  
PRIOR APPLICATION NUMBER: 60/106248  
PRIOR FILING DATE: 1998-10-29  
PRIOR APPLICATION NUMBER: 60/106464  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: 60/106905  
PRIOR FILING DATE: 1998-11-03  
PRIOR APPLICATION NUMBER: 60/108787  
PRIOR FILING DATE: 1998-11-17  
PRIOR APPLICATION NUMBER: 60/108801  
PRIOR FILING DATE: 1998-11-17  
PRIOR APPLICATION NUMBER: 60/108849  
PRIOR FILING DATE: 1998-11-18  
PRIOR APPLICATION NUMBER: 60/112422  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113296  
PRIOR FILING DATE: 1998-12-22

PRIOR APPLICATION NUMBER: 60/113605  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/113621  
PRIOR FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: 60/115558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115565  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115733  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119549  
PRIOR FILING DATE: 1999-02-10  
PRIOR APPLICATION NUMBER: 60/123518  
PRIOR FILING DATE: 1999-03-10  
PRIOR APPLICATION NUMBER: 60/125259  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 60/125775  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 60/126773  
PRIOR FILING DATE: 1999-03-29  
PRIOR APPLICATION NUMBER: 60/127887  
PRIOR FILING DATE: 1999-04-05  
PRIOR APPLICATION NUMBER: 60/130232  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/131022  
PRIOR FILING DATE: 1999-04-26  
PRIOR APPLICATION NUMBER: 60/131270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131291  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/131445  
PRIOR FILING DATE: 1999-04-28  
PRIOR APPLICATION NUMBER: 60/134287  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/140650  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/140723  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/141037  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/144758  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/145698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 60/146222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: 60/146963  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/149320  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/149638  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/151733  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: 60/164418  
PRIOR FILING DATE: 1999-11-09  
PRIOR APPLICATION NUMBER: 60/166361  
PRIOR FILING DATE: 1999-11-16  
PRIOR APPLICATION NUMBER: 60/169445  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: 60/169495  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: 60/169835

Query Match 65.5%; Score 1110.5; DB 9; Length 667;  
Best Local Similarity 76.6%; Pred. No. 1,6e-91;  
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

CY 1 CYSALVLSCLTFVLMKSLVTHR-----LGSQSGSD-----GQFSWNLR 42  
DB 365 CYSALVLSCLTFVLMKSLVTHRTNLRALHFGALDLSPLRSPHPSRQALFCWMSFS 424  
CY 43 VPLPPLAGLVQOIIIFLGTTLALFLVLMPLVHLGNLULFFNSLSSWPFMTLLAVI 102

```
Db 425 AVGTAFICGLVQIIIFFLGTTALAFVLMPLVHGNLLPFSLESSNPFMTTALAIV 484
Qy 103 LONMAAHVWFLETHDGHQOLNRRVLYAATFLLPPLNVIVGAMXXXC----- 149
Db 485 LONMAAHVWFLETHDGHQOLNRRVLYAATFLLPPLNVIVGAMVATRWVLSALYNAIHL 544
Qy 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSPAMTAFCSLLLOAOSLLPRTMAAPOD 203
Db 545 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSPAMTAFCSLLLOAOSLLPRTMAAPOD 604
Qy 204 SLRPEDEGMOQLLOTQDSMAKARPGAKRGARAGLATTLLHNTLQVFRKTALLGANG 263
Db 605 SLRPEDEGMOQLLOTQDSMAKARPGAKRGARAGLATTLLHNTLQVFRKTALLGANG 664
Qy 264 AQP 266
Db 665 AQP 667
```

## RESULT 6

```
US-10-218-631-80
/ Sequence 80, Application US/10218631
/ Publication No. US20030045687A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Gerritsen, Mary
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe F.
/ APPLICANT: Watanabe, Colin L.
/ APPLICANT: Wood, William I.
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3530P1C14
/ CURRENT APPLICATION NUMBER: US/10/218,631
/ PRIOR FILING DATE: 2002-08-12
/ PRIOR APPLICATION NUMBER: 10/119,480
/ PRIOR FILING DATE: 2002-04-09
/ PRIOR APPLICATION NUMBER: 60/059113
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/062287
/ PRIOR FILING DATE: 1997-10-17
/ PRIOR APPLICATION NUMBER: 60/063549
/ PRIOR FILING DATE: 1997-10-28
/ PRIOR APPLICATION NUMBER: 60/064103
/ PRIOR FILING DATE: 1997-10-31
/ PRIOR APPLICATION NUMBER: 60/069873
/ PRIOR FILING DATE: 1997-12-17
/ PRIOR APPLICATION NUMBER: 60/078910
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/079294
/ PRIOR FILING DATE: 1998-03-25
/ PRIOR APPLICATION NUMBER: 60/079656
/ PRIOR FILING DATE: 1998-03-26
/ PRIOR APPLICATION NUMBER: 60/079728
/ PRIOR FILING DATE: 1998-03-27
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 246
/ SEQ ID NO 80
/ LENGTH: 667
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-218-631-80
```

```
Query Match 65.5% Score 1110.5; DB 9; Length 667;
Best Local Similarity 76.6%; Pred. No. 1.6e-91;
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;
```

```
Qy 1 CYSALVLSCLTFLVLMRSLVTHR-----LGSQGSQD-----GQFSWNLFS 42
Db 365 CYSALVLSCLTFLVLMRSLVTHRNLALHRAALDLSPLHRSPPSPQALFCWMSFS 424
Qy 43 VPLPLPPLAGLAVQOIIIFFLGTTALAFVLMPLVHGNLLPFSLESSNPFMTTALAIV 102
Db 425 AVGTAFICGLVQIIIFFLGTTALAFVLMPLVHGNLLPFSLESSNPFMTTALAIV 484
Qy 103 LONMAAHVWFLETHDGHQOLNRRVLYAATFLLPPLNVIVGAMXXXC----- 149
Db 485 LONMAAHVWFLETHDGHQOLNRRVLYAATFLLPPLNVIVGAMVATRWVLSALYNAIHL 544
Qy 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSPAMTAFCSLLLOAOSLLPRTMAAPOD 203
Db 545 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSPAMTAFCSLLLOAOSLLPRTMAAPOD 604
Qy 204 SLRPEDEGMOQLLOTQDSMAKARPGAKRGARAGLATTLLHNTLQVFRKTALLGANG 263
Db 605 SLRPEDEGMOQLLOTQDSMAKARPGAKRGARAGLATTLLHNTLQVFRKTALLGANG 664
Qy 264 AQP 266
Db 665 AQP 667
```

## RESULT 7

```
US-10-230-338-80
/ Sequence 80, Application US/10230338
/ Publication No. US20030044934A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Gerritsen, Mary
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe F.
/ APPLICANT: Watanabe, Colin L.
/ APPLICANT: Wood, William I.
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3530P1C2
/ CURRENT APPLICATION NUMBER: US/10/230,338
/ PRIOR FILING DATE: 2002-08-28
/ PRIOR APPLICATION NUMBER: 10/119,480
/ PRIOR FILING DATE: 2002-04-09
/ PRIOR APPLICATION NUMBER: 60/059113
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/062287
/ PRIOR FILING DATE: 1997-10-17
/ PRIOR APPLICATION NUMBER: 60/063549
/ PRIOR FILING DATE: 1997-10-28
/ PRIOR APPLICATION NUMBER: 60/064103
/ PRIOR FILING DATE: 1997-10-31
/ PRIOR APPLICATION NUMBER: 60/069873
/ PRIOR FILING DATE: 1997-12-17
/ PRIOR APPLICATION NUMBER: 60/078910
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/079294
/ PRIOR FILING DATE: 1998-03-25
/ PRIOR APPLICATION NUMBER: 60/079656
/ PRIOR FILING DATE: 1998-03-26
/ PRIOR APPLICATION NUMBER: 60/079728
/ PRIOR FILING DATE: 1998-03-27
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 246
/ SEQ ID NO 80
/ LENGTH: 667
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-230-338-80
```

Query Match 65.5%; Score 110.5; DB 9; Length 667;  
Best Local Similarity 76.6%; Pred. No. 1.6e-91;  
Matches 231; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFLVLMRSIVTHR-----LGSQSGSD-----GQFSWNIFS 42  
DB 365 CYSALVLSCLTFLVLMRSIVTHR-----LGSQSGSD-----GQFSWNIFS 424  
QY 43 VPLPLPPLAGLLVQOIIFFLGTTALAFVLMVHGRNLLPFRSLESSWPFMTLALAVI 102  
DB 425 AVOTAFICIGLLVQOIIFFLGTTALAFVLMVHGRNLLPFRSLESSWPFMTLALAVI 484  
QY 103 LQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149  
DB 485 LQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 544  
QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMTAFCSLLLOAQSLPRTMAAPD 203  
DB 545 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMTAFCSLLLOAQSLPRTMAAPD 604  
QY 204 SLRGEDEGMOQLLOTQDSMAKARGARPGARWGLAYTLNHPVLQVFRKTLALGANG 263  
DB 605 SLRGEDEGMOQLLOTQDSMAKARGARPGARWGLAYTLNHPVLQVFRKTLALGANG 664  
QY 264 AQP 266  
DB 665 AQP 667

RESULT 8  
US-09-759-056-5  
Sequence 5, Application US/09759056  
Patent No. US20020156252A1  
GENERAL INFORMATION:  
APPLICANT: Pennica, Diane  
APPLICANT: Smith, Victoria  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: No. US20020156252A1 STRA6 Polypeptides  
FILE REFERENCE: GENENT 2827A2  
CURRENT APPLICATION NUMBER: US/09/759,056  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/197089  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/175849  
PRIOR FILING DATE: 2000-01-13  
PRIOR APPLICATION NUMBER: 60/228914  
PRIOR FILING DATE: 2000-08-29  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 658  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-759-056-5

Query Match 65.3%; Score 110.5; DB 9; Length 658;  
Best Local Similarity 76.2%; Pred. No. 3.5e-91;  
Matches 231; Conservative 5; Mismatches 30; Indels 37; Gaps 4;

DB 536 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMTAFCSLLLOAQSLPRTMAAPD 595  
QY 204 SLRGEDEGMOQLLOTQDSMAKARGARPGARWGLAYTLNHPVLQVFRKTLALGANG 263  
DB 596 SLRGEDEGMOQLLOTQDSMAKARGARPGARWGLAYTLNHPVLQVFRKTLALGANG 655  
QY 264 AQP 266  
DB 656 AQP 658

RESULT 9  
US-09-901-812-5  
Sequence 5, Application US/09901812  
Patent No. US20020173461A1  
GENERAL INFORMATION:  
APPLICANT: Pennica, Diane  
APPLICANT: Polakis, Paul  
APPLICANT: Szeto, Wayne  
TITLE OF INVENTION: UPREGULATION OF TUMOR ANTIGENS TO  
FILE REFERENCE: GENENT 083A  
CURRENT APPLICATION NUMBER: US/09/901,812  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/228,914  
PRIOR FILING DATE: 2000-08-29  
PRIOR APPLICATION NUMBER: 09/759,056  
PRIOR FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: 60/175,849  
PRIOR FILING DATE: 2000-01-13  
PRIOR APPLICATION NUMBER: 60/197,089  
PRIOR FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 658  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-901-812-5

Query Match 65.3%; Score 110.5; DB 9; Length 658;  
Best Local Similarity 76.2%; Pred. No. 3.5e-91;  
Matches 231; Conservative 5; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFLVLMRSIVTHR-----LGSQSGSD-----GQFSWNIFS 42  
DB 365 CYSALVLSCLTFLVLMRSIVTHR-----LGSQSGSD-----GQFSWNIFS 415  
QY 43 VPLPLPPLAGLLVQOIIFFLGTTALAFVLMVHGRNLLPFRSLESSWPFMTLALAVI 102  
DB 416 AVOTAFICIGLLVQOIIFFLGTTALAFVLMVHGRNLLPFRSLESSWPFMTLALAVI 475  
QY 103 LQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 149  
DB 476 LQNMMAHWFLETHDGHQPOLTNRRVLYAATFLLPPLNVLVGAMXXXC----- 535  
QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMTAFCSLLLOAQSLPRTMAAPD 203  
DB 536 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMTAFCSLLLOAQSLPRTMAAPD 595  
QY 204 SLRGEDEGMOQLLOTQDSMAKARGARPGARWGLAYTLNHPVLQVFRKTLALGANG 263  
DB 596 SLRGEDEGMOQLLOTQDSMAKARGARPGARWGLAYTLNHPVLQVFRKTLALGANG 655  
QY 264 AQP 266  
DB 656 AQP 658

RESULT 10  
US-09-816-653a-7  
Sequence 7, Application US/0981653A

Publication No. US20030021788A1  
 GENERAL INFORMATION:  
 APPLICANT: Raetelli, Luca  
 TITLE OF INVENTION: NOVEL HUMAN STRA6-LIKE PROTEIN AND NUCLEIC ACIDS ENCODING THE SAME  
 FILE REFERENCE: 10716/57  
 CURRENT APPLICATION NUMBER: US/09/816,653A  
 PRIOR FILING DATE: 2001-03-23  
 PRIOR APPLICATION NUMBER: 60/191,532  
 PRIOR FILING DATE: 2000-03-23  
 NUMBER OF SEQ ID NOS: 7  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 7  
 LENGTH: 670  
 TYPE: PRT  
 ORGANISM: Mus musculus retinoic acid-responsive protein (Stra6) mRNA, complete cds  
 US-09-816-653A-7

Query Match 49.1%; Score 812.5; DB 9; Length 670;  
 Best Local Similarity 57.7%; Pred. No. 1,2e-66;

Matches 176; Conservative 28; Mismatches 62; Indels 39; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLYTHR-----LQSGGSGD-----GQFSNNLFS 42  
 DB 366 CYISALVLSCLTFLVLMRSVLYTHR-----LQSGGSGD-----GQFSNNLFS 425  
 QY 43 VPLPLPPLAGLVVOQIIFFLGTALAFVLMVPLHGRNLLFRSLSSWPPMLTLALAVI 102  
 DB 426 AYQTAASCLGLLVQOVIFFLGTALAFVLMVPLHGRNLLFRSLSSWPPMLTLALAVI 485  
 QY 103 LONMAAHVFLFETHDGPOLTNRRVLYAATFLLPLNVLVGAMXXXSPSIAIRHPT-- 159  
 DB 486 LONIAAHVFLFETHDGPOLTNRRVLYAATFLLPLNVLVGAMXXXSPSIAIRHPT-- 545  
 QY 160 -----PGYYTRNFKTEVSGSHPMATFCSLLLOQSLPTPTMAAPD 203  
 DB 546 GQMDLSLTPORASLDPGYHTYONFLRIEASGSHGVAFACALLHAPSQPPPLAPD 605  
 QY 204 SLRPGEDBGMQLQTKYDAMAKAPGAXRGRARWGLAYTLNHPILQVRKTXALLG--A 261  
 DB 606 SLRPAEEDBGMQLQTKYDAMAKAGKGSOSRARKGLAYTLNHPISLQAFRKALTSAKA 665  
 QY 262 NGAQP 266  
 DB 666 NGTQP 670

RESULT 11  
 US-09-925-299-796  
 Sequence 796, Application US/09925299  
 Publication No. US20030040617A9  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 FILE REFERENCE: PA102  
 CURRENT APPLICATION NUMBER: US/09/925,299  
 PRIOR FILING DATE: 2001-08-10  
 PRIOR APPLICATION NUMBER: PCT/US00/05883  
 PRIOR FILING DATE: 2000-03-08  
 PRIOR APPLICATION NUMBER: 60/124,270  
 NUMBER OF SEQ ID NOS: 1556  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 796  
 LENGTH: 178  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-925-299-796

Query Match 28.6%; Score 485; DB 9; Length 178;  
 Best Local Similarity 88.2%; Pred. No. 3.3e-36;  
 Matches 97; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 36 FSNLFSVPLPLPPLAGLVVOQIIFFLGTALAFVLMVPLHGRNLLFRSLSSWPPML 95  
 DB 27 FCMSPSAVOTAFICLGLVVOIIFLGTALAFVLMVPLHGRNLLFRSLSSWPPML 86  
 QY 96 TLALAVIIONMAAHVFLFETHDGPOLTNRRVLYAATFLLPLNVLVGAM 145  
 DB 87 TLALAVIIONMAAHVFLFETHDGPOLTNRRVLYAATFLLPLNVLVGAM 136

RESULT 12  
 US-09-925-299-796  
 Sequence 796, Application US/09925299  
 Patent No. US2002005627A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 FILE REFERENCE: PA102  
 CURRENT APPLICATION NUMBER: US/09/925,299  
 PRIOR FILING DATE: 2001-08-10  
 PRIOR APPLICATION NUMBER: PCT/US00/05883  
 PRIOR FILING DATE: 2000-03-08  
 PRIOR APPLICATION NUMBER: 60/124,270  
 NUMBER OF SEQ ID NOS: 1556  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 796  
 LENGTH: 178  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-925-299-796

Query Match 28.6%; Score 485; DB 10; Length 178;  
 Best Local Similarity 88.2%; Pred. No. 3.3e-36;  
 Matches 97; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 36 FSNLFSVPLPLPPLAGLVVOQIIFFLGTALAFVLMVPLHGRNLLFRSLSSWPPML 95  
 DB 27 FCMSPSAVOTAFICLGLVVOIIFLGTALAFVLMVPLHGRNLLFRSLSSWPPML 86  
 QY 96 TLALAVIIONMAAHVFLFETHDGPOLTNRRVLYAATFLLPLNVLVGAM 145  
 DB 87 TLALAVIIONMAAHVFLFETHDGPOLTNRRVLYAATFLLPLNVLVGAM 136

RESULT 13  
 US-09-826-508-30  
 Sequence 30, Application US/09826508  
 Patent No. US20010025099A1  
 GENERAL INFORMATION:  
 APPLICANT: Nabil Elshourbagy  
 TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides  
 FILE REFERENCE: GP-70744USB  
 CURRENT APPLICATION NUMBER: US/09/826,508  
 PRIOR FILING DATE: 2001-04-05  
 NUMBER OF SEQ ID NOS: 40  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 30  
 LENGTH: 403  
 TYPE: PRT  
 ORGANISM: HOMO SAPIENS  
 US-09-826-508-30

Query Match 5.8%; Score 97.5; DB 10; Length 403;  
 Best Local Similarity 19.9%; Pred. No. 0.46;  
 Matches 69; Conservative 30; Mismatches 107; Indels 141; Gaps 14;

QY 5 ALVSCSLTFLVLMRSVLYTHRSLGSGGQFSNNLFSVPLPLPPLAGLVVOQII----- 59  
 DB 134 ALVSCSLTFLVLMRSVLYTHRSLGSGGQFSNNLFSVPLPLPPLAGLVVOQII----- 181  
 QY 60 -----FFLGTTALAFVLMVPLHGRNLLFRS 86



```

Db 182 VLTVDTRPACAYEPMDFVVALIYDVLVLTGLALFTLCKFKRMKINGAFLLITAF 241
Qy 87 LE-SSMPFWLT-----ALAVILQMAAHVPLETHDGHPOLTNRRVL 128
Db 242 LSVLIWVAMMTYVLFNGVKLOOGDAMNDPTLAILT--AASGVFVIFH-AIPEI----- 292
Qy 129 YAATFLLEPLNVLVGAMXXXCPSIAIRHPTPGYYTYRNFLKIEVSQSHPMAMTAFCSLL 188
Db 293 -----HCTLLPALQENTPNYF-----DTSQPRMRETAFA----- 320
Qy 189 QAOSLLPRTMAAPQDSLPRGEDEBGMOLLQTKDSMAKAGPAXKRGARWGLAYTLLNHP 248
Db 321 EEDVOLPRAYM-----ENKAFSMDEHNALRTAGFPNGLSK----- 357
Qy 249 TLOVFRKTTALGANGAOPCSSLPGSPPSITPAMO-----PAGPPDHXG 291
Db 358 -----RPSGSLGKRPSAPFRSNVYOPTEMAVVLNGGTTPTAPPSHTG 399

```

```

RESULT 14
US-09-895-686-5
; Sequence 5, Application US/09895686
; Patent No. US20020106655A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Tang, Y. Tom
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN GPCR PROTEINS
; FILE REFERENCE: PC-0044 CIP
; CURRENT APPLICATION NUMBER: US/09/895,686
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020106655A1 2705201CD1
US-09-895-686-5

```

```

Query Match 5.8%; Score 97.5; DB 10; Length 403;
Best Local Similarity 19.9%; Pred. No. 0.46;
Matches 69; Conservative 30; Mismatches 107; Indels 14; Gaps 14;

```

```

Qy 5 ALVLSGLTFLVLMRSIVTHRIKSGSGDGFNMNLSVPLPLPLAGLLVQOII----- 59
Db 134 ALCFSCSLSQAMRVRLVRH-----GTGPRAGMOLVGLALCL-----MLVQYIIVAEWL 181
Qy 60 -----FLGTTALAFVLMP-----VLHGRNLLFFRS 86
Db 182 VLTVDTRPACAYEPMDFVVALIYDVLVLTGLALFTLCKFKRMKINGAFLLITAF 241
Qy 87 LE-SSMPFWLT-----ALAVILQMAAHVPLETHDGHPOLTNRRVL 128
Db 242 LSVLIWVAMMTYVLFNGVKLOOGDAMNDPTLAILT--AASGVFVIFH-AIPEI----- 292
Qy 129 YAATFLLEPLNVLVGAMXXXCPSIAIRHPTPGYYTYRNFLKIEVSQSHPMAMTAFCSLL 188
Db 293 -----HCTLLPALQENTPNYF-----DTSQPRMRETAFA----- 320
Qy 189 QAOSLLPRTMAAPQDSLPRGEDEBGMOLLQTKDSMAKAGPAXKRGARWGLAYTLLNHP 248
Db 321 EEDVOLPRAYM-----ENKAFSMDEHNALRTAGFPNGLSK----- 357
Qy 249 TLOVFRKTTALGANGAOPCSSLPGSPPSITPAMO-----PAGPPDHXG 291
Db 358 -----RPSGSLGKRPSAPFRSNVYOPTEMAVVLNGGTTPTAPPSHTG 399

```

```

RESULT 15
US-09-826-508-32
; Sequence 32, Application US/09826508
; Patent No. US20010025099A1
; GENERAL INFORMATION:
; APPLICANT: Nabli Elshoutbagy
; APPLICANT: Lisa Vawter
; TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides
; FILE REFERENCE: GP-70744USB
; CURRENT APPLICATION NUMBER: US/09/826,508
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 427
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-826-508-32

```

```

Query Match 5.8%; Score 97.5; DB 10; Length 427;
Best Local Similarity 19.9%; Pred. No. 0.49;
Matches 69; Conservative 30; Mismatches 107; Indels 14; Gaps 14;

```

```

Qy 5 ALVLSGLTFLVLMRSIVTHRIKSGSGDGFNMNLSVPLPLPLAGLLVQOII----- 59
Db 158 ALCFSCSLSQAMRVRLVRH-----GTGPRAGMOLVGLALCL-----MLVQYIIVAEWL 205
Qy 60 -----FLGTTALAFVLMP-----VLHGRNLLFFRS 86
Db 206 VLTVDTRPACAYEPMDFVVALIYDVLVLTGLALFTLCKFKRMKINGAFLLITAF 265
Qy 87 LE-SSMPFWLT-----ALAVILQMAAHVPLETHDGHPOLTNRRVL 128
Db 266 LSVLIWVAMMTYVLFNGVKLOOGDAMNDPTLAILT--AASGVFVIFH-AIPEI----- 316
Qy 129 YAATFLLEPLNVLVGAMXXXCPSIAIRHPTPGYYTYRNFLKIEVSQSHPMAMTAFCSLL 188
Db 317 -----HCTLLPALQENTPNYF-----DTSQPRMRETAFA----- 344
Qy 189 QAOSLLPRTMAAPQDSLPRGEDEBGMOLLQTKDSMAKAGPAXKRGARWGLAYTLLNHP 248
Db 345 EEDVOLPRAYM-----ENKAFSMDEHNALRTAGFPNGLSK----- 381
Qy 249 TLOVFRKTTALGANGAOPCSSLPGSPPSITPAMO-----PAGPPDHXG 291
Db 382 -----RPSGSLGKRPSAPFRSNVYOPTEMAVVLNGGTTPTAPPSHTG 423

```

```

Search completed: March 13, 2003, 11:37:59
Job time : 15.645 secs

```



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:27:42 ; Search time 16.7462 Seconds

(without alignments)  
571.023 Million cell updates/sec

Title: US-09-816-653A-4

Perfect score: 1695  
Sequence: 1 CYSALVLSGLTFLVLMRS.....APPTWMSQGFWRKLVG 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfillseq.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	93	5.5	2016	4	US-09-634-920-4
2	87	5.1	618	3	US-08-866-381A-6
3	87	5.1	621	1	US-08-208-887A-49
4	87	5.1	621	4	US-09-280-598-18
5	87	5.1	621	4	US-08-945-771-4
6	83	4.9	539	4	US-09-036-887A-11
7	83	4.9	539	4	US-09-370-700-11
8	83	4.9	1495	4	US-08-462-467B-12
9	82.5	4.9	1024	4	US-09-562-737-84
10	82.5	4.9	2972	4	US-08-469-260A-387
11	82.5	4.9	5087	4	US-09-144-085-1
12	81.5	4.8	376	2	US-08-875-972-2
13	81.5	4.8	6095	4	US-09-144-085-2
14	81	4.8	410	3	US-09-083-521-2
15	81	4.8	419	1	US-08-439-131A-3
16	81	4.8	419	1	US-08-440-674-2
17	80	4.7	859	1	US-08-395-580-2
18	80	4.7	859	5	PCT-US95-02792-2
19	80	4.7	2618	4	US-09-413-814-28
20	79.5	4.7	2588	3	US-08-936-135-2
21	79	4.7	235	1	US-07-940-605A-12
22	79	4.7	235	2	US-08-680-096-12
23	79	4.7	1306	4	US-08-999-774A-13
24	79	4.7	1385	2	US-08-667-399-7
25	77.5	4.6	376	3	US-08-751-512-8
26	77.5	4.6	869	1	US-08-188-582-32
27	77.5	4.6	869	1	US-08-646-715-32

28	77	4.5	448	1	US-08-207-904-2	Sequence 2, Appli
29	77	4.5	448	1	US-08-207-904-17	Sequence 17, Appl
30	77	4.5	1017	4	US-09-600-776-6	Sequence 6, Appli
31	76.5	4.5	355	1	US-07-759-568-1	Sequence 1, Appli
32	76.5	4.5	355	1	US-08-450-393A-8	Sequence 8, Appli
33	76.5	4.5	355	2	US-08-390-000A-5	Sequence 5, Appli
34	76.5	4.5	355	4	US-08-446-669-8	Sequence 8, Appli
35	76.5	4.5	355	5	PCT-US95-00476-8	Sequence 8, Appli
36	76.5	4.5	360	1	US-08-202-056-7	Sequence 7, Appli
37	76	4.5	194	4	US-08-959-004-9	Sequence 9, Appli
38	76	4.5	195	4	US-09-042-071-41	Sequence 41, Appl
39	76	4.5	350	1	US-08-458-067-2	Sequence 2, Appli
40	76	4.5	350	5	PCT-US96-07795-2	Sequence 2, Appli
41	76	4.5	350	5	PCT-US96-07796-2	Sequence 2, Appli
42	76	4.5	635	1	US-07-832-855-2	Sequence 2, Appli
43	76	4.5	635	4	US-08-176-320-2	Sequence 2, Appli
44	76	4.5	724	4	US-08-914-999-4	Sequence 4, Appli
45	75.5	4.5	197	3	US-09-010-809-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-09-634-920-4  
Sequence 4, Application US/09634920  
Patent No. 6342357  
GENERAL INFORMATION:  
APPLICANT: Splawski, Igor  
TITLE OF INVENTION: Keating, Mark T  
TITLE OF INVENTION: ALTERATIONS IN THE LONG QT SYNDROME GENES KVLQT1 AND  
FILE REFERENCE: 2323-155  
CURRENT APPLICATION NUMBER: US/09/634,920  
CURRENT FILING DATE: 2000-08-09  
PRIOR APPLICATION NUMBER: 60/190,057  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/147,488  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 2016  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-634-920-4  
Query Match 5.5%; Score 93; DB 4; Length 2016;  
Best Local Similarity 23.8%; Pred. No. 0.85;  
Matches 67; Conservative 31; Mismatches 113; Indels 70; Gaps 13;  
QY 35 QRSNNLFSVPLPLPLAGLVQOIIIFLIGTTALAFLVMVLYLGRNLLFFRSLSSEWPM 94  
DB 778 QQQNNIFD-----SIVTSLMEIGLSRMSNL---SVLSFLLVFKLAASWPTL 825  
QY 95 -----LTLAVALILQNMMAAHV-----FLETHDGHPOLYTR---RVLY 129  
DB 826 NTLIKIGNSVGLNLTVLVLAITVIFFAVVGQMLFQKNVSELRDSDGLPRMHMDF 885  
QY 130 AATFLPLPLNVGVGMKXCCPSIAIRHPTRGYTTNPLKIVSOSHPRMTRFCSLLIQ 189  
DB 886 HAFLIIF--RIICGMIWTMDCMVSGQS---LCILVFLVNVINLVNLFALLLS 940  
QY 190 AQSILPRTAAQDSLRGEDEGQVLQOTKDSMAKCARPGARCGARVGLAVTLL--HNP 248  
DB 941 SFS-----ADULTAPDDBRENNQL--ALARIQRLRVRVKRTTWDFCCGLLRHRP 989  
QY 249 TLQVFRKTRLLANGAOP--CSSLPQSPSITPAMQAPGPPD 288  
DB 990 -----QKPAALAAQQLFSCIATPYSP-----PPE 1015  
RESULT 2

US-08-866-381A-6  
Sequence 6, Application US/08866381A  
Patent No. 6045797  
GENERAL INFORMATION:  
APPLICANT: Ben Lewis Margolis  
APPLICANT: Joseph Schlesinger  
TITLE OF INVENTION: METHODS FOR TREATMENT OR DIAGNOSIS  
TITLE OF INVENTION: OF DISEASES OR CONDITIONS ASSOCIATED  
TITLE OF INVENTION: WITH A BLM DOMAIN  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: Storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: PASTESEO for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/866,381A  
FILING DATE: May 30, 1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/212,234  
FILING DATE: March 14, 1994  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 226/043  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEO ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 618 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
OTHER INFORMATION: GRS-10  
US-08-866-381A-6

Query Match 5.1%; Score 87; DB 3; Length 618;  
Best Local Similarity 21.7%; Pred. No. 0.75;  
Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;

QY 172 EVSOSHPTAMTAFCSL-----LLO-----AQSLLPRTMAAPDLSRPGEDSG 213  
DB 1 DINSVSLSNACNMQSDTTPALPLEDQHASNOGAASSSRGQPAQSPKXORSQPHI 60  
QY 214 MOLLQTKDSMAKCARPGAXRGARAGLAVTLIHNPTLOVPRKTKALLGANGAPCCSLPGS 273  
DB 61 LRLQEDDQLRTA-----SLPAIPNPFPE-----LTGA-----APGS 93  
QY 274 PSITPAM---OPAGP-DHXGXVEVCJHMEPXXGSAF-PTWLWESQOGFWRKXL 323  
DB 94 PPSVAPSSLPPEPSQPAHGCRCCKWIPGENTRNGKRIKRWQPPGFQLSKL 148

RESULT 3  
US-08-208-887A-49  
Sequence 49, Application US/08208887A  
Patent No. 5677421

GENERAL INFORMATION:  
APPLICANT: Schlesinger, Joseph  
APPLICANT: Skolnick, Edward Y.  
APPLICANT: Margolis, Benjamin L.  
TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR  
TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: 10036-2711  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/208,887A  
FILING DATE: 11-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7683-063  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEO ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 621 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-208-887A-49

Query Match 5.1%; Score 87; DB 1; Length 621;  
Best Local Similarity 21.7%; Pred. No. 0.75;  
Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;

QY 172 EVSOSHPTAMTAFCSL-----LLO-----AQSLLPRTMAAPDLSRPGEDSG 213  
DB 4 DINSVSLSNACNMQSDTTPALPLEDQHASNOGAASSSRGQPAQSPKXORSQPHI 63  
QY 214 MOLLQTKDSMAKCARPGAXRGARAGLAVTLIHNPTLOVPRKTKALLGANGAPCCSLPGS 273  
DB 64 LRLQEDDQLRTA-----SLPAIPNPFPE-----LTGA-----APGS 96  
QY 274 PSITPAM---OPAGP-DHXGXVEVCJHMEPXXGSAF-PTWLWESQOGFWRKXL 323  
DB 97 PPSVAPSSLPPEPSQPAHGCRCCKWIPGENTRNGKRIKRWQPPGFQLSKL 151

RESULT 4  
US-09-280-598-18  
Sequence 18, Application US/09280598  
Patent No. 6391584  
GENERAL INFORMATION:  
APPLICANT: Schlesinger, Joseph  
APPLICANT: Skolnick, Edward Y.  
APPLICANT: Margolis, Benjamin L.  
APPLICANT: App. Harold  
TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR  
TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas

CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/280,598  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/252,820  
 FILING DATE: 02-JUN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Coruzzi, Laura A.  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 7683-067  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 621 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 US-09-280-598-18

Query Match 5.1%; Score 87; DB 4; Length 621;  
 Best Local Similarity 21.7%; Pred. No. 0.75;  
 Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;

QY 172 EVSOSHPIAMTAFCSL-----LIQ-----AQSILPRTMAPODSLPGEDG 213  
 DB 4 DINSVESINSACNMOSDTPAPLEDDGHSNQAASSSRGQPSPRKMKOROPVHI 63  
 QY 214 MQLLOTDSMAKARGARAGARAGLAATLHNFTLQVFRKTKALLGANAQPCSSLPSS 273  
 DB 64 LRRLOEDDQQLRFA-----SLPAIPNPPE-----LTGA-----APGS 96  
 QY 274 PPSITPAM---QPAAPP-DHXGKVEVCLHMEPYXGSAF-PTWLMSQOGFWRKLT 323  
 DB 97 PPSVAPSSLPPEPSQPPAHGCRCKWIPGENTRNGRKXIMRWQPPFGQLSKL 151

RESULT 5  
 US-08-945-771-4  
 Sequence 4, Application US/08945771  
 Patent No. 6465623  
 GENERAL INFORMATION:  
 APPLICANT: Daly, Roger J.  
 APPLICANT: Sutherland, Robert L.  
 TITLE OF INVENTION: GDU, A novel signalling protein  
 FILE REFERENCE: 273402001700  
 CURRENT APPLICATION NUMBER: US/08/945,771  
 CURRENT FILING DATE: 1998-04-22  
 EARLIER APPLICATION NUMBER: PCT/US96/00258  
 EARLIER FILING DATE: 1996-MAY-02  
 NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 4  
 LENGTH: 621  
 TYPE: PRT  
 ORGANISM: Mus musculus  
 US-08-945-771-4

Query Match 5.1%; Score 87; DB 4; Length 621;  
 Best Local Similarity 21.7%; Pred. No. 0.75;  
 Matches 38; Conservative 28; Mismatches 59; Indels 50; Gaps 8;

QY 172 EVSOSHPIAMTAFCSL-----LIQ-----AQSILPRTMAPODSLPGEDG 213  
 DB 4 DINSVESINSACNMOSDTPAPLEDDGHSNQAASSSRGQPSPRKMKOROPVHI 63  
 QY 214 MQLLOTDSMAKARGARAGARAGLAATLHNFTLQVFRKTKALLGANAQPCSSLPSS 273  
 DB 64 LRRLOEDDQQLRFA-----SLPAIPNPPE-----LTGA-----APGS 96  
 QY 274 PPSITPAM---QPAAPP-DHXGKVEVCLHMEPYXGSAF-PTWLMSQOGFWRKLT 323  
 DB 97 PPSVAPSSLPPEPSQPPAHGCRCKWIPGENTRNGRKXIMRWQPPFGQLSKL 151

RESULT 6  
 US-09-036-987A-11  
 Sequence 11, Application US/09036987A  
 Patent No. 6143526  
 GENERAL INFORMATION:  
 APPLICANT: Baltz, Richard H.  
 APPLICANT: Broughton, Mary C.  
 APPLICANT: Crawford, Kathryn P.  
 APPLICANT: Madduri, Krishnamurthy  
 APPLICANT: Merlo, Donald J.  
 APPLICANT: Treadway, Patti J.  
 APPLICANT: Turner, Jan R.  
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
 TITLE OF INVENTION: Production  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Dow Agrosciences LLC Patent Department  
 STREET: 9330 Zionsville Road  
 CITY: Indianapolis  
 STATE: Indiana  
 COUNTRY: USA  
 ZIP: 46268  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/036,987A  
 FILING DATE: 09-MAR-1998  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stuart, Donald R.  
 REGISTRATION NUMBER: 28,479  
 REFERENCE/DOCKET NUMBER: 50,608  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (317)337-4816  
 TELEFAX: (317)337-4847  
 INFORMATION FOR SEQ ID NO: 11:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 539 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-036-987A-11

Query Match 4.9%; Score 83; DB 4; Length 539;  
 Best Local Similarity 19.9%; Pred. No. 1.7;  
 Matches 62; Conservative 38; Mismatches 122; Indels 90; Gaps 16;

QY 30 GSGDQGS-----W-----NLFSVP-----LPLPLAGLLVQOITTFGLTALAFLVLMPTL 76  
 DB 217 GGGGNGGIYTFWLTPTDVPVSTDAEMLPRPPTVLL----- 254  
 QY 77 HGRNLFFRSLESSNPFWLTLAVIIONMAAHVVFETHDGHPOULTNRVLYAATFLFL 136  
 DB 255 -----RSFHPHHELTQSFVLLQNF-GNN--YEQGSA----- 285

QY 137 PLNVLPAMXXXXSPSIALRHPTTPYUYYRFLKLEIVSGSHRMTAFCSLLLOAQSLPR 196  
 Db 286 PESTQLQFS----TLVCANRQAGUYV---LNVHLDGTDPR--NAERTLAENLSAIAAQ 334  
 QY 197 TMAAPODSLRPGSEEDGMQLQTKDSNAKGARPCAXHGRARWGAVATLLHNPTL-QVPRK 255  
 Db 335 VGVTPRAQGR--ETLPMLRSTQVAGALIEGGEPRMORTKKKAAUYLRGTGISEAQATVYRR 392  
 QY 256 TALLGANGAOPCSGL-----PGSPRSITPAMQAPGRPDHXGVYECILM-EPXGSAAPT 309  
 Db 393 LTVYGYVD--NPAAALLLLLGCGMANAVAPSBTALQARDVSLKALPVTNMSEPADEHNLT 450  
 QY 310 WLWESQOGFMRK 321  
 Db 451 WI---RGFYRE 458

```

RESULT 7
US-09-370-700-11
: Sequence 11 Application US/09370700
: Patent No. 6274350
: GENERAL INFORMATION:
: APPLICANT: Baltz, Richard R
: APPLICANT: Broughton, Mary C
: APPLICANT: Crawford, Kathryn P
: APPLICANT: Madduri, Krishnamurthy
: APPLICANT: Treadway, Patrick
: APPLICANT: Turner, Jan R
: APPLICANT: Waldron, Gilve
: TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticides
: FILE REFERENCE: 50489 Div1
: CURRENT APPLICATION NUMBER: US/09/370,700
: EARLIER FILING DATE: 1999-08-09
: EARLIER APPLICATION NUMBER: US 09/36987
: NUMBER OF SEQ ID NOS: 39
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 11
: LENGTH: 539
: TYPE: PR1
: ORGANISM: Saccharopolyspora spinosa
: US-09-370-700-11

```

Query Match 4.9%; Score 83; DB 4; Length 539;  
Best Local Similarity 19.9%; Pred. NO. 1.7;  
Matches 62; Conservative 38; Mismatches 122; Indels 90; Gaps 16

Qy	137	P L V L V G A M X X X C S P S I A I R P T P G Y T Y Y N F L K I E V S O G H P M A T A F C S L L L Q A O S L I P R	196
Db	286	P E S T Q L G L S ----- T L V C A H R Q A Q Y V T ----- L N V H L D G T D P - N A E R T L A E H L S A I N A Q	334
Qy	197	T M A P O D S L R P G E D E G M O L L O T K D S M A K G A R F A X H G R A R M G L A V T L L H N P T L - O V R K	255
Db	335	V G V T P A E G R R - E T L P M L R S T Q V A G A I E G E G P E M Q K T K Y K A I L R T G S E M Q L T Y Y R K	392
Qy	256	T A L L G A N G A O P C S S I ----- P C S P S I T P M Q P A R P P D H K G X V E C L H K - E P X Y G S A P T	309
Db	393	L T Y T Y G T D - N P A A L L L L G Y G M A N A V A P S K T A L A O R D S V L K A L P V T W S P A E D E R H L T	450
Qy	310	W L M E S O O G F W R K	321
Db	451	W I ----- R G F Y R E	458

## RESULT 8

```

US-08-462-467B-12
Sequence 12 Application US/08462467B
Patent No. 6210899
GENERAL INFORMATION:
APPLICANT: Rosenbaum, Jan S
TITLE OF INVENTION: The Use of a BMP Protein Receptor
TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
TITLE OF INVENTION: Co-Transfected with a Type II BMP Receptor and a Type I
TITLE OF INVENTION: BMP Receptor
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSER: The Procter & Gamble Company
STREET: 11810 East Miami River Road
CITY: Ross
STATE: OH
COUNTRY: USA
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,467B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hersko, Bart S.
REGISTRATION NUMBER: 32,572
REFERENCE/DOCKET NUMBER: 5474R
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 627-0633
TELEFAX: (513) 627-0260
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1495 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-462-467B-12

```

Query Match 4.9%; Score 83; DB 4; Length 1495;  
 Best Local Similarity 24.4%; Pred. NO. 6.6;  
 Matches 74; Conservative 32; Mismatches 117; Indels 80; Gaps 13

QY	11	-----	GGNNLFFRSLESSSPWMLTALAVLQNMMAHWVFETHDGHQPLTNRRVLY	129
Db	1213	YRSERPHEGLYVEILEIILETR-PEIMETALAA--RGARGCYSILETRG-----LY	1260	
QY	130	AATFLPLPLNVLVGAMXXXCPSIAIRHPTPGYYTR--NFKLIEVSQSHAMTAFCSLL	167	
Db	1261	G-----LYILEVALGGLTYRGSLNDEPR--YTRTYRASMETVALRGRSRAP-----	1306	
QY	168	LQAGSLLPPT-----MAAPDQSR-----PGEEDGAGNLOLT--	219	
Db	1307	-----PSEETTYGGLASPMETIRRGVALVALVALCYALLYSKRGLEARGPILVALS	1355	
QY	220	-KSMKAGKAPKAGKARGKGLAYTLILHNPITOVRKTKALLGANGAOPCSSLGFSPSIT	278	
Db	1360	ERSSNNAGTRPMSNSERASPCLCYSLEARGALAV---ALLEIYSLEMETSERGCLCYSTR	1415	
QY	279	PAM	281	
Db	1416	PAL	1418	

## RESULT 9

US-09-562-737-84

Sequence 84, Application US/09562737  
 Patent No. 6428967  
 GENERAL INFORMATION:  
 APPLICANT: Herz, Joachim  
 APPLICANT: Gotthardt, Michael  
 TITLE OF INVENTION: LDL Receptor Signaling Pathways  
 FILE REFERENCE: UTSW0708  
 CURRENT APPLICATION NUMBER: US/09/562,737  
 CURRENT FILING DATE: 2000-05-01  
 NUMBER OF SEQ ID NOS: 132  
 SOFTWARE: Patentin Ver. 2.1  
 SEQ ID NO 84  
 LENGTH: 1024  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 OTHER INFORMATION: Sequence  
 US-09-562-737-84

Query Match 4.9%; Score 82.5; DB 4; Length 1024;  
 Best Local Similarity 19.8%; Pred. No. 4.5;  
 Matches 51; Conservative 36; Mismatches 85; Indels 85; Gaps 10;

QY 35 QFSNLSVPLPLPLAGLIVQIIFPLGTALAFVLYLMPVHGRNLFPSLESSMPFW 94  
 DB 772 QENWNIFD-----GIYQLSLMELG---LARVEGLSVLRSSRLRVPKLTATSP-T 818  
 QY 95 LITALLAVILQNMAMHVFLETHDGHPLQTNRRVLYATFLLPLNV--VGANXXXXCSFS 152  
 DB 819 LNNLIYIIGNSVGLM-----NLTVLAIIFYLFAVVGQAFGKSKYKCCVDK 865  
 QY 153 IAIRHPTP-----GYTYR-----NFIKLEVSQSHP 178  
 DB 866 INVDCLEPMHNDPFGSLIYFRLCGHWIETMMDCMIVAGTWCILKFMALMVGNNV 925  
 QY 179 AMTAFCSLLQQLPRTMAAPQSLRPGEDSGMQLQTKXSMAGAPGAXRGRAH 238  
 DB 926 VNLFLFALLNLSSFS-----SDNLQATDDDNEMNLQ-----IAGVRMOS 964  
 QY 239 GLAYTLHNPTLQVFRK 255  
 DB 965 GIDF--VRKKITCEFRK 979

RESULT 10  
 US-08-469-260A-387  
 Sequence 387, Application US/08469260A  
 Patent No. 6451578  
 GENERAL INFORMATION:  
 APPLICANT: JOHN N. SIMONS  
 APPLICANT: TAMM J. PILOT-MATIAS  
 APPLICANT: GEORGE J. DAMSON  
 APPLICANT: GEORGE G. SCHLAUDER  
 APPLICANT: SURESH M. DESAI  
 APPLICANT: THOMAS P. LEARY  
 APPLICANT: ANTHONY SCOTT MUEHROFF  
 APPLICANT: JAMES C. ERKER  
 APPLICANT: SHERI L. BUIJK  
 APPLICANT: ISA K. MUSHAMMAR  
 TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
 TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
 NUMBER OF SEQUENCES: 716  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: ABBOTT LABORATORIES D377/APED  
 STREET: 100 ABBOTT PARK ROAD  
 CITY: ABBOTT PARK  
 STATE: IL  
 COUNTRY: USA  
 ZIP: 60064-3500  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/469,260A  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/424,550  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: POROMBSKI, PRISCILLA E.  
 REGISTRATION NUMBER: 33,207  
 REFERENCE/DOCKET NUMBER: 5527.PC.01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 708-937-6365  
 TELEFAX: 708-938-2623  
 INFORMATION FOR SEQ ID NO: 387:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2972 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULAR TYPE: protein  
 US-08-469-260A-387

Query Match 4.9%; Score 82.5; DB 4; Length 2972;  
 Best Local Similarity 24.8%; Pred. No. 19;  
 Matches 50; Conservative 14; Mismatches 75; Indels 63; Gaps 9;

QY 132 TELFLPLNVVGAMXXXXSPSIAIRPPTGYTTRNFKLEV--SQSPANTAFCSLLQA 190  
 DB 2800 TILMYP-----RHPIRYVLLIPHYLMAVYRGSTPDELVNCVQGNH 2841  
 QY 191 QS-----LPRTMAA-----PODSLRPCEB-----DEGMQLQTKSMAGAPCA 231  
 DB 2442 YSFPLPLPRVLVSLHGPWCLOVTDSITKTRMEGSKLRDGMKSLAMHRRRGVTRTR 2901  
 QY 232 XRGRAWG-LATYLLHNPTLQVFRKTLGANGAPCSSLPGPSSTTPMAGAPPDHX 290  
 DB 2902 LRGGKXWGLRALRLMXPKX-----EXXPINSILPG-----FOLATPEYEH 2943  
 QY 291 GXVEVCLHNSPXXSAPPTLM 312  
 DB 2944 EYVLISIK-----SRPPIRW 2959

RESULT 11  
 US-09-144-085-1  
 Sequence 1, Application US/09144085  
 Patent No. 6280999  
 GENERAL INFORMATION:  
 APPLICANT: Gustafson, Claes  
 APPLICANT: Bastiaen, Mary C.  
 APPLICANT: Ashley, Gary  
 APPLICANT: Jullien, Bryan  
 APPLICANT: Ziemann, Rainer  
 TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
 TITLE OF INVENTION: THEREFOR  
 FILE REFERENCE: 30062-20020, 20  
 CURRENT APPLICATION NUMBER: US/09/144,085  
 CURRENT FILING DATE: 1998-08-31  
 EARLIER APPLICATION NUMBER: 09/010,809  
 EARLIER FILING DATE: 1998-01-22  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: Patentin Ver. 2.0  
 SEQ ID NO 1  
 LENGTH: 5087  
 TYPE: PRT  
 ORGANISM: Sorangium cellulosum  
 US-09-144-085-1  
 Query Match 4.9%; Score 82.5; DB 4; Length 5087;  
 Best Local Similarity 29.2%; Pred. No. 39;  
 Matches 33; Conservative 12; Mismatches 47; Indels 21; Gaps 5;

QY 143 GAKXXCSPSIAIRHPTPGYTYRNFLKIEVSQHPA--MTAFCSLLQAQ-SILPRTMA 199  
 DB 234 GALAEVFGPERKARP-----LYGSSKSNLGHTPAGVGVKWLMSQHEVLPTTH 288  
 QY 200 APQDSLRFGEDEGMOLOTKDSMAKAPGAXRGAR-----WGLATYLLH 246  
 DB 289 AEQPSPHIGWEGSLDGE-----ARPMRRNGRARARAGVSSFGISGTNAH 334

## RESULT 12

US-08-875-972-2  
 / Sequence 2, Application US/08875972  
 / Patent No. 5985564  
 / GENERAL INFORMATION:  
 / APPLICANT: Huntington Potter and Jinhue Li  
 / TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING  
 / NUMBER OF SEQUENCES: 29  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
 / STREET: Two Millita Drive  
 / CITY: Lexington  
 / STATE: Massachusetts  
 / COUNTRY: USA  
 / ZIP: 02173-4799  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / OPERATING SYSTEM: IBM PC compatible  
 / SOFTWARE: PC-DOS/MS-DOS  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/875,972  
 / FILING DATE: 08-AUG-97  
 / CLASSIFICATION: 435  
 / PRIORITY APPLICATION DATA:  
 / APPLICATION NUMBER: US 60/002,448  
 / FILING DATE: 16-AUG-1995  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Granahan Bag, Patricia  
 / REGISTRATION NUMBER: 32,227  
 / REFERENCE/DOCKET NUMBER: H095-03PA  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (781) 861-6240  
 / TELEFAX: (781) 861-9540  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 376 amino acids  
 / TYPE: amino acid  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / US-08-875-972-2

Query Match 4.8%; Score 81.5; DB 2; Length 376;  
 Best Local Similarity 20.1%; Pred. No. 1.5;  
 Matches 65; Conservative 35; Mismatches 114; Indels 109; Gaps 13;

QY 1 CY-----ISALVLSCLTFLVLRSLVTHRLS-----GSGGDQGFQSWNL 40  
 DB 98 CYKPIHGMIMSSLMFLFTYIYLGSEVUKITVVMNDYPTLLLVVMFGAVGCHVIMWK- 156  
 QY 41 FSVPLPLPLAGLLVQOIIFFLGTALAFVLMVJHGRNLLFRSLSSWPFVLTAA-- 98  
 DB 157 -----GFLVEQAVLIMISA-----LMP-----LMFIKYPPEWSAVVILAPI 193  
 QY 99 -----LAVIILONMAAHVVFLETHDGHQPLTNRRVLYAA-----TELLFPLN-VLVGAKMXXC 149  
 DB 194 SVYDLVTVLCSSTPLKMLVETAOERNETTFSPLITSSPMVWTVMNSKIDPESQALQLPY 253  
 QY 150 SPSTAIRHPTPGYTYRNFLKIEVSQHPAMTAFCSLLQAQSLPRTMAAPQDSLRFGE 209  
 DB 254 DPEWEDSYDSFCSPSYPEVFE-----PLIAGYPCGELEEEF 289

QY 210 EDEG--MOLOTKDSMAKAPGAXRGARW-----GLATYLLHPTLOVFRK 255  
 DB 290 ESQGVKVLVIGTSTVYVPGGAPPTSGDWITTLACFVALIGLCITLL--LLAVFK 346  
 QY 256 TALIGANGAQCSSLFGEPSPT 278  
 DB 347 -----ALPALPISIT 356

## RESULT 13

US-09-144-085-2  
 / Sequence 2, Application US/09144085  
 / Patent No. 6280999  
 / GENERAL INFORMATION:  
 / APPLICANT: Gustafsson, Claes  
 / APPLICANT: Betlach, Mary C.  
 / APPLICANT: Ashley, Gary  
 / APPLICANT: Julien, Bryan  
 / APPLICANT: Ziemann, Rainer  
 / TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
 / FILE REFERENCE: 30062-20020.20  
 / CURRENT APPLICATION NUMBER: US/09/144,085  
 / CURRENT FILING DATE: 1998-08-31  
 / EARLIER APPLICATION NUMBER: 09/010,809  
 / EARLIER FILING DATE: 1998-01-22  
 / NUMBER OF SEQ ID NOS: 8  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 2  
 / LENGTH: 6095  
 / TYPE: PR1  
 / ORGANISM: Sorangium cellulosum  
 / US-09-144-085-2

Query Match 4.8%; Score 81.5; DB 4; Length 6095;  
 Best Local Similarity 29.2%; Pred. No. 64;  
 Matches 33; Conservative 12; Mismatches 47; Indels 21; Gaps 5;

QY 143 GAKXXCSPSIAIRHPTPGYTYRNFLKIEVSQHPA--MTAFCSLLQAQ-SILPRTMA 199  
 DB 352 GALVEVFGPERKARP-----LYGSSKSNLGHGAPGAVGVKWLMSQHEVLPTTH 406  
 QY 200 APQDSLRFGEDEGMOLOTKDSMAKAPGAXRGAR-----WGLATYLLH 246  
 DB 407 AEQPSPHIGWEGSLDGE-----ARPMRRNGRARARAGVSSFGISGTNAH 452

## RESULT 14

US-09-083-521-2  
 / Sequence 2, Application US/09083521  
 / Patent No. 6046970  
 / GENERAL INFORMATION:  
 / APPLICANT: Lal, Preeti  
 / APPLICANT: Guegler, Karl J.  
 / APPLICANT: Corley, Neil C.  
 / TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS  
 / NUMBER OF SEQUENCES: 7  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 / STREET: 3174 PORTER DRIVE  
 / CITY: PALO ALTO  
 / STATE: CALIFORNIA  
 / COUNTRY: USA  
 / ZIP: 94304  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / OPERATING SYSTEM: IBM PC compatible  
 / SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/09/083,521  
 / FILING DATE: Herewith  
 / CLASSIFICATION:







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 11:25:47 ; Search time 40.9351 Seconds  
(without alignments)  
1057.923 Million cell updates/sec

Title: US-09-816-653a-4  
Perfect score: 1695  
Sequence: 1 CYISALVLSCLTFLVLMRS.....APPTLMESOGFMRKILVG 325

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq.101002:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1134	66.9	660	23	AAU91277
2	1110.5	65.5	560	22	AAB94108
3	1110.5	65.5	667	22	AAU04557
4	1110.5	65.5	667	22	AAB88572
5	1110.5	65.5	667	23	AAU78575
6	1110.5	65.5	667	23	AAU83631
7	1110.5	65.5	667	23	AAB05048
8	1110.5	65.5	669	23	AAU91278
9	1106.5	65.3	658	22	AAU04558
10	1106.5	65.3	658	23	AAU78576

11	1060.5	62.6	653	23	AAB05046	Human NOV8a protei
12	1060.5	62.6	247	22	AAU23581	Human NOV8a protei
13	1002.5	59.1	247	22	AAU23581	Murine EST encoded
14	485	28.6	178	21	AAB53256	Human colon cancer
15	485	28.6	178	21	AAG74057	Human colon cancer
16	479	28.3	149	23	AAB89878	Human polypeptide
17	440	26.0	88	20	AAW88559	Secreted protein e
18	440	26.0	88	22	AAB50326	Human secreted pro
19	115	6.8	487	23	AAU17374	Human retinoic aci
20	97.5	5.8	403	20	AAU32141	Human G-protein co
21	97.5	5.8	403	21	AAU32141	Human ORFX ORF2849
22	97.5	5.8	403	21	AAU57287	Human GPCR protein
23	97.5	5.8	403	21	AAU78615	Human protein SEO
24	97.5	5.8	403	22	AAU93548	Human polypeptide
25	97.5	5.8	403	23	AAB81627	Human GPCR-like
26	97.5	5.8	427	20	AAU32142	Human G-protein co
27	95.5	5.6	526	22	AAU91958	C glutamylcum prote
28	95	5.6	476	22	AAU52307	Human TRIP6. Homo
29	95	5.6	618	16	AAB80165	Mouse signal trans
30	93.5	5.5	1603	22	AAU19518	Human diagnostic a
31	93	5.5	2015	22	AAB82242	Human SCNSA mutat
32	93	5.5	2016	19	AAU23594	Human hhl sodium c
33	93	5.5	2016	22	AAB82239	Human SCNSA protei
34	93	5.5	2016	22	AAB82240	Human SCNSA mutat
35	93	5.5	2016	22	AAB82241	Human SCNSA mutat
36	93	5.5	2016	22	AAB82243	Human SCNSA mutat
37	93	5.5	2016	22	AAB82244	Human SCNSA mutat
38	93	5.5	2016	22	AAB82245	Human SCNSA mutat
39	92	5.4	664	22	AAB87308	ABC transport rela
40	92	5.4	2144	23	AAU09174	Human transporter
41	92	5.4	2146	22	AAU04463	Human PD-ATP-Bind
42	92	5.4	2180	23	AAU04463	Human transporter
43	91.5	5.4	572	23	AAU76036	Human sugar transp
44	90	5.3	361	22	AAB10502	Novel human diagno
45	90	5.3	477	22	AAU03472	Human gene 7 encod

#### ALIGNMENTS

RESULT 1	AAU91277	standard; Protein: 660 AA.
ID	AAU91277	
XX	AAU91277;	
AC	18-JUN-2002 (first entry)	
XX		
DT		
XX		
DE	Human NOV2a protein.	
XX		
KW	Human: NOVX; gene therapy; cardiomyopathy; atherosclerosis;	
KW	diabetes; cell signal processing; metabolic pathway modulation;	
KW	inflammation; autoimmune disorder; scleroderma; transplantation;	
KW	allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;	
KW	graft versus host disease; Leisch-Nyhan syndrome; peridontitis;	
KW	pancreatitis; musculoskeletal disorder; Parkinson's disease;	
KW	Huntington's disease; behavioural disorder; pain; obesity; wound healing;	
KW	neurodegenerative disorder; neuropsychiatric disorder; hypertension;	
KW	growth disorder; reproductive disorder; lung disease.	
XX		
OS	Homo sapiens.	
XX		
FN	WO200216600-A2.	
PD	28-FEB-2002.	
XX		
PF	27-AUG-2001; 2001WO-US26518.	
XX		
PR	25-AUG-2000; 2000US-227800P.	
PR	25-AUG-2000; 2000US-228305P.	
PR	25-AUG-2000; 2000US-228324P.	
PR	30-AUG-2000; 2000US-228397P.	
PR	30-AUG-2000; 2000US-229185P.	



CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH0166 to AAH1328 and  
 CC AAH1363 to AAH18742 represent human cDNA sequences. AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 560 AA;

SO Query Match 65.54; Score 1110.5; DB 22; Length 560;

Best Local Similarity 76.64; Pred. No. 1.1e-106;

Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFLVLRSLVTHR-----LGSGSGSD-----GQFSWMLFS 42  
 DB 258 CYSALVLSCLTFLVLRSLVTHRNRALHRAALDLPRLSRPHPSQAIFCWMSPS 317  
 QY 43 VPLPPLAGLLVVOIIFPLGTALAPVLMPTVHGNNLFPRLSSSPFWTLALAVI 102  
 DB 318 AVGTAFICGLLVQOIIFPLGTALAPVLMPTVHGNNLFPRLSSSPFWTLALAVI 377  
 QY 103 LQNMAMHWFLETHGHGHPOLTRRVLVYATFLPPLNVLYGAMXXC----- 149  
 DB 378 LQNMAMHWFLETHGHGHPOLTRRVLVYATFLPPLNVLYGAMVATRWLSALVNAIHL 437  
 QY 150 -SPSIAIRHP-----TPGYTYRNPFLKTESQSHPMATFQSLILQOSILPRTMAPOD 203  
 DB 438 GQMDLSILPPRAATLDPGYTYRNPFLKTESQSHPMATFQSLILQOSILPRTMAPOD 497  
 QY 204 SLRPGDEGMOLLQTKOSMACAPGAXRGRARWGLAYTLTHNPTLOVERKTKALIGANG 263  
 DB 498 SLRPGDEGMOLLQTKOSMACAPGAXRGRARWGLAYTLTHNPTLOVERKTKALIGANG 557  
 QY 264 AQP 266  
 DB 558 AQP 560

# RESULT 3

ID AAU04557 standard; Protein; 667 AA.

XX AAU04557;

XX 26-SEP-2001 (first entry)

DE Human Str66 homologue, PRO10282.

XX Human; Str66; PRO10282; 15q23; tumour; breast cancer;  
 KM head tumour; leukaemia; lymphoid malignancy; inflammatory disorder;  
 KW immunogenic disorder; antigen; antibody.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 8..13 /note= "Asn is N-glycosylated"

FT Region 50..56 /label= N\_myristoylation\_site

FT Domain 54..69 /label= Transmembrane\_domain

FT Region 89..97 /note= "Region deleted in Str66 variant PRO19578"

FT Domain 102..119 /label= Transmembrane\_domain

FT Region 132..140 /note= "ATP/GTP-binding site motif A (P-loop)"

FT Domain 148..166 /label= Transmembrane\_domain

FT Region 176..182 /label= N\_myristoylation\_site

FT Domain 207..222 /label= Transmembrane\_domain  
 FT Region 241..247 /label= N\_myristoylation\_site  
 FT Domain 301..320 /label= Transmembrane\_domain  
 FT Region 317..323 /label= N\_myristoylation\_site  
 FT Region 341..347 /label= N\_myristoylation\_site  
 FT Region 364..375 /label= N\_myristoylation\_site  
 FT Region 364..375 /note= "Prokaryotic membrane lipoprotein attachment site"  
 FT Domain 364..380 /label= Transmembrane\_domain  
 FT Domain 431..451 /label= Transmembrane\_domain  
 FT Domain 474..489 /label= Transmembrane\_domain  
 FT Region 525..531 /label= N\_myristoylation\_site  
 FT Domain 560..535 /label= Transmembrane\_domain  
 FT Region 627..633 /label= N\_myristoylation\_site  
 FT Region 631..637 /label= N\_myristoylation\_site  
 FT Region 640..646 /label= N\_myristoylation\_site  
 FT Region 661..667 /label= N\_myristoylation\_site

PN MO200151635-A2.

PD 19-JUL-2001.

XX 11-JAN-2001; 2001WO-US00847.

XX 13-JAN-2000; 2000US-0175849.

XX 14-APR-2000; 2000US-0197089.

XX 29-AUG-2000; 2000US-0228914.

XX (GETH ) GENENTECH INC.

XX Pennica D, Smith V, Wood WT;

XX WPI; 2001-442146/47.

XX N-PSDB; AAS08630.

XX Human nucleic acid encoding a PRO10282 or PRO19578 polypeptide (the PRO

PT polypeptides are referred to as Str66 polypeptides), useful in

PT molecular biology, including use as hybridisation probes, and in

XX chromosome and gene mapping -

XX Claim 24; Fig 2; 159pp; English.

CC The sequence is human PRO10282, a homologue of murine Str66, a  
 CC retinoic acid responsive protein thought to play an important role in  
 CC early dorsoventral limb patterning during development and later in the  
 CC control of endochondral ossification. The gene for the Str66 homologue  
 CC is located on chromosome 15q23. The Str66 polypeptides, agonists,  
 CC antagonists or anti-Str66 antibodies are useful for preparation of a  
 CC medicament useful in the treatment of a condition which is responsive to  
 CC the Str66 polypeptides, agonists, antagonists or anti-Str66 antibodies.  
 CC The Str66 polypeptides may also be employed as molecular weight markers  
 CC for protein electrophoresis. The Str66 nucleic acids have applications in  
 CC molecular biology, including use as hybridisation probes, and in  
 CC chromosome and gene mapping. The antibodies and other anti-tumour  
 CC compounds may be used to treat various conditions, including those  
 CC characterised by overexpression and/or activation of amplified genes.  
 CC Exemplary conditions or disorders include benign or malignant tumours  
 CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,  
 CC colorectal, prostate, pancreatic, lung, vulval, thyroid, hepatic

CC Carcinomas, sarcomas, glioblastomas, and various head and neck tumours,  
 CC leukemias and lymphoid malignancies), neuronal, glial, astrocytal,  
 CC hypothalamic and other glandular, macrophagal, epithelial, stromal and  
 CC blastococic disorders, inflammatory, angiogenic and immunogenic  
 CC disorders.

XX Sequence 667 AA;

Query Match 65.5%; Score 1110.5; DB 22; Length 667;  
 Best Local Similarity 76.6%; Pred. No. 1.4e-106;  
 Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLYTHR-----LGSQSGSD-----GQFSNVLFS 42  
 Db 365 CYISALVLSCLTFLVLMRSVLYTHRNLALHGAALDLPRLSPHPSRQALFCWMSFS 424  
 QY 43 VPLPLPPLAGLVQOIIFFLGTALAFVLMVPLHGNLLFFRSLESSWPFWTLALAVI 102  
 Db 425 AVGTAFICGLLVQOIIFFLGTALAFVLMVPLHGNLLFFRSLESSWPFWTLALAVI 484  
 QY 103 LQNMAMHWFLETHDGHQPLTNRRVLYAATFLPLVNVLYGAMXXXC----- 149  
 Db 485 LQNMAMHWFLETHDGHQPLTNRRVLYAATFLPLVNVLYGAMVATWRVLSALYNAILH 544  
 QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATPFCSLLOAQSLPRTMAAPD 203  
 Db 545 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMATPFCSLLOAQSLPRTMAAPD 604  
 QY 204 SLRGEDEGMQLLOTKDSMAKARPGARAGWGLAYTLNHPLOVERKTALLGANG 263  
 Db 605 SLRGEDEGMQLLOTKDSMAKARPGARAGWGLAYTLNHPLOVERKTALLGANG 664  
 QY 264 AQP 266  
 ||||  
 Db 665 AQP 667

RESULT 4  
 AAB8572  
 ID AAB8572 standard; Protein; 667 AA.

XX AAB8572;

DT 04-JUN-2001 (first entry)

XX Human hydrophobic domain containing protein clone HP10713 #36.

XX Human; hydrophobic domain; immunosuppressant; anti-HIV; neuroprotective;  
 XX antianaemic; vlnnary; antilucer; osteoporotic; anti-inflammatory;  
 XX cytostatic; gene therapy; autoimmune disorder; multiple sclerosis;  
 XX HIV infection; anemia; burn; ulcer; osteoporosis; tumour; wound healing;  
 XX inflammatory bowel disease; nutritional supplement; appetite; vaccine;  
 XX behavioural characteristic; immune response.

OS Homo sapiens.

PN W0200112660-A2.

XX 22-FEB-2001.

PF 10-AUG-2000; 2000W0-JP05356.

XX 17-AUG-1999; 99QP-0230344.

PR 07-SEP-1999; 99QP-0252551.

PR 01-OCT-1999; 99QP-0281132.

PR 22-OCT-1999; 99QP-0301624.

PR 04-NOV-1999; 99QP-0313877.

XX (SAGA ) SAGAMI CHEM RES CENT.  
 PA (PROT-) PROTEGENE INC.  
 XX Kato S, Kimura T;  
 PI  
 XX

DR WPI, 2001-160059/16.  
 DR N-PSDB; AAF94442.

XX Human proteins with hydrophobic domains and the DNAs which encode them  
 PT are useful for treating autoimmune disorders, burns and tumors and for  
 PT screening novel pharmaceuticals -

PS Claim 1; Page 288-291; 51bpb; English.

CC AAF94417 to AAF94516 encode the human proteins given in AAB8572 to  
 CC AAB8606 (1) which have a hydrophobic domain. (1) have immunosuppressant,  
 CC anti-HIV, neuroprotective, antianaemic, vlnnary, antilucer,  
 CC osteopathic, anti-inflammatory and cytostatic activities, and can be  
 CC used in gene therapy. (1) can be used as pharmaceuticals and as antigens  
 CC to prepare antibodies. DNA and cDNA (II) encoding (1) can be used as  
 CC probes for genetic diagnosis and gene sources for gene therapy or for  
 CC producing (1) in large quantities. Cells containing (1) are used for  
 CC the detection of ligands or receptors corresponding to membrane or  
 CC secretory proteins and to screen small molecule novel pharmaceuticals.  
 CC Antibodies directed to (1) can be used for the detection, quantification  
 CC and purification of (1). Activities of (1) may include cytokine and cell  
 CC proliferation/differentiation function, immune stimulating or suppressing  
 CC activity, haematopoiesis regulating activity, tissue growth activity,  
 CC activity/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity and anti-inflammatory  
 CC activity. (1) and (II) can be used to treat autoimmune disorders e.g.  
 CC multiple sclerosis, HIV infections, anaemia, burns, ulcers, osteoporosis,  
 CC inflammatory bowel disease and tumours. (1) and (II) can also be used for  
 CC wound healing, as nutritional sources or supplements e.g. as amino acid,  
 CC carbon or nitrogen source, to effect metabolism, catabolism, anabolism,  
 CC processing and utilisation of dietary fat, protein, carbohydrate,  
 CC vitamins and minerals, to effect behavioural characteristics, to affect  
 CC appetite, and can act as antigens in vaccinee to raise an immune response  
 CC to the protein or another material cross-reactive with the protein.

XX Sequence 667 AA;

Query Match 65.5%; Score 1110.5; DB 22; Length 667;  
 Best Local Similarity 76.6%; Pred. No. 1.4e-106;  
 Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLMRSVLYTHR-----LGSQSGSD-----GQFSNVLFS 42  
 Db 365 CYISALVLSCLTFLVLMRSVLYTHRNLALHGAALDLPRLSPHPSRQALFCWMSFS 424  
 QY 43 VPLPLPPLAGLVQOIIFFLGTALAFVLMVPLHGNLLFFRSLESSWPFWTLALAVI 102  
 Db 425 AVGTAFICGLLVQOIIFFLGTALAFVLMVPLHGNLLFFRSLESSWPFWTLALAVI 484  
 QY 103 LQNMAMHWFLETHDGHQPLTNRRVLYAATFLPLVNVLYGAMXXXC----- 149  
 Db 485 LQNMAMHWFLETHDGHQPLTNRRVLYAATFLPLVNVLYGAMVATWRVLSALYNAILH 544  
 QY 150 -SPSIAIRHP-----TPGYTYRNFLKIEVSQSHPMATPFCSLLOAQSLPRTMAAPD 203  
 Db 545 GQMDLSLPPRAATLDPGYTYRNFLKIEVSQSHPMATPFCSLLOAQSLPRTMAAPD 604  
 QY 204 SLRGEDEGMQLLOTKDSMAKARPGARAGWGLAYTLNHPLOVERKTALLGANG 263  
 Db 605 SLRGEDEGMQLLOTKDSMAKARPGARAGWGLAYTLNHPLOVERKTALLGANG 664  
 QY 264 AQP 266  
 ||||  
 Db 665 AQP 667

RESULT 5  
 AAU78575  
 ID AAU78575 standard; Protein; 667 AA.

XX AAU78575;

DT 18-JUN-2002 (first entry)

XX Human Strag6 (PRO10282) protein.  
 DE Human; cancer; Strag6; PRO10282; cytostatic; stromal disorder;  
 XX tumor; retinoid; Mt; tumour antigen; leukaemia; lymphoid malignancy;  
 KM human; retinoid; Mt; tumour antigen; leukaemia; lymphoid malignancy;  
 KM neuronal disorder; glial disorder; astrocytoma; glioma;  
 KM hypothalamic disorder; glandular disorder; microphagal disorder;  
 KM epithelial disorder; blastocoele disorder; chromosome 15q23.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Region 8..12  
 FT /note= "N Glycosylation site"  
 FT Region 50..56  
 FT /note= "N-myristoylation site"  
 FT Domain 54..69  
 FT /note= "Transmembrane domain"  
 FT Domain 102..119  
 FT /note= "Transmembrane domain"  
 FT Domain 132..140  
 FT /note= "Transmembrane domain"  
 FT Region 148..166  
 FT /note= "Arp/GTP binding site motif A"  
 FT Domain 176..182  
 FT /note= "Transmembrane domain"  
 FT Region 207..222  
 FT /note= "N-myristoylation site"  
 FT Domain 241..247  
 FT /note= "Transmembrane domain"  
 FT Region 301..320  
 FT /note= "N-myristoylation site"  
 FT Domain 317..323  
 FT /note= "Transmembrane domain"  
 FT Region 341..347  
 FT /note= "N-myristoylation site"  
 FT Domain 364..380  
 FT /note= "N-myristoylation site"  
 FT Domain 364..380  
 FT /note= "Transmembrane domain"  
 FT Region 364..375  
 FT /note= "Transmembrane domain"  
 FT Region 431..451  
 FT /note= "lipoprotein lipid attachment site"  
 FT Domain 474..489  
 FT /note= "Transmembrane domain"  
 FT Domain 525..531  
 FT /note= "Transmembrane domain"  
 FT Region 560..535  
 FT /note= "N-myristoylation site"  
 FT Domain 627..633  
 FT /note= "Transmembrane domain"  
 FT Region 631..637  
 FT /note= "N-myristoylation site"  
 FT Region 640..646  
 FT /note= "N-myristoylation site"  
 FT Region 661..667  
 FT /note= "N-myristoylation site"  
 FT Region 661..667  
 FT /note= "N-myristoylation site"  
 EN WO200218608-A2.  
 XX  
 XX 07-MAR-2002.  
 PD  
 PF 10-JUL-2001; 2001MO-0821635.  
 XX  
 XX 29-AUG-2000; 2000US-228914P.  
 PR 11-JAN-2001; 2001US-0759056.  
 PR 10-JUL-2001; 2001US-0901612.  
 XX  
 XX (GENTECH ) GENENTECH INC.  
 XX  
 XX Tice D, Pennica D, Polakis P, Szeto W;  
 PI  
 XX WPI; 2002-292202/33.  
 DR  
 DR N-PSDB; ABK47278.

XX Selectively enhancing the expression of a protein in a tumour cell  
 PT characterised by aberrant Wnt signalling, useful for enhancing the  
 PT efficacy of treatment of cancers, comprises treating the tumour cell  
 PT with a retinoid  
 XX  
 XX Example 1; Fig 2; 141pp; English.  
 PS  
 XX This invention relates to a novel method for selective enhancement of  
 CC the expression of a protein in a tumour cell characterised by aberrant  
 CC Wnt signalling by treating the tumour cell with a retinoid. The method  
 CC is useful for enhancing the efficacy of treatment of cancers  
 CC characterised by aberrant Wnt signalling, and for identifying tumour  
 CC antigens the expression of which is selectively enhanced by retinoid  
 CC treatment. The method is also useful in the early detection of  
 CC tumours for early intervention. The tumour antigens are useful as  
 CC targets for cancer therapy, since selective enhancement of their  
 CC expression by retinoid treatment relative to normal cells improves the  
 CC efficacy and therapeutic index of cancer therapeutics directed against  
 CC these antigens. Antibodies binding to the tumour antigens may be  
 CC administered for the treatment of various disorders or conditions,  
 CC including those characterised by overexpression and/or activation of the  
 CC tumour antigens, where such conditions or disorders include benign or  
 CC malignant tumours; leukaemias and lymphoid malignancies; neuronal, glial,  
 CC astrocytal, hypothalamic and other glandular, microphagal, epithelial,  
 CC stromal and blastocoele disorders. The present sequence represents  
 CC the human Strag6 protein (PRO10282) which is selectively upregulated by  
 CC retinoid using the method of the invention. The gene encoding the human  
 CC Strag6 protein is located on human chromosome 15q23.  
 CC  
 XX  
 XX Sequence 667 AA;  
 SQ  
 Query Match 65.5%; Score 110.5; DB 23; Length 667;  
 Best Local Similarity 76.6%; Pred. No. 1.4e-106;  
 Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;  
 QY 1 CYSALVLSCLTFLVLRSLVTHR-----LGSGGSD-----GQFSNMIFS 42  
 DB 365 CYSALVLSCLTFLVLRSLVTHRSLRHRGALDLSPLHRSPRQALFCMMS 424  
 QY 43 VPLPLPPLAGLIVQIIFFLGTTALAPLVLPVLRGNLFFRSLESSWPFMTLAVI 102  
 DB 425 AVQTAPICGLIVQIIFFLGTTALAPLVLPVLRGNLFFRSLESSWPFMTLAVI 484  
 QY 103 LQNMAMWVFLSTHGHQPLTRRVLYATFLFPLVWVGAKXKX----- 149  
 DB 485 LQNMAMWVFLSTHGHQPLTRRVLYATFLFPLVWVGAMVATWRVLSALVNAHL 544  
 QY 150 -SPSIAIRP-----TPGYTYRNFLKIEVSQSHPAMTAPCSLLAQSLPRTMAAPD 203  
 DB 545 GQMDLSLPRBATLDPGYTYRNFLKIEVSQSHPAMTAPCSLLAQSLPRTMAAPD 604  
 QY 204 SLRPGEDSGMQLQTKDSMAKAPGAXRGRAGVLAATLNPLOVERKTLGANG 263  
 DB 605 SLRPGEDSGMQLQTKDSMAKAPGAXRGRAGVLAATLNPLOVERKTLGANG 664  
 QY 264 AQP 266  
 DB 665 AQP 667  
 DB  
 XX  
 XX RESULT 6  
 XX AAU83631  
 XX ID AAU83631 standard; Protein; 667 AA.  
 XX  
 XX AAU83631;  
 XX  
 XX 08-MAY-2002 (first entry)  
 XX  
 XX Human PRO protein, Seq ID No 80.  
 DE  
 XX Human; secreted protein; PRO; tumour; lung cancer; colon cancer;  
 XX breast cancer; prostate tumour; rectal tumour; liver tumour;  
 KW

KM pericyte cell proliferation; chondrocyte cell proliferation;  
 KM tumour necrosis factor-alpha.  
 XX Homo sapiens.  
 OS WO200208288-A2.  
 XX  
 XX 31-JAN-2002.  
 PD  
 XX 29-JUN-2001; 2001WO-US21066.  
 PF  
 XX 20-JUL-2000; 2000US-219556P.  
 PR 25-JUL-2000; 2000US-220585P.  
 PR 25-JUL-2000; 2000US-220605P.  
 PR 25-JUL-2000; 2000US-220607P.  
 PR 25-JUL-2000; 2000US-220624P.  
 PR 25-JUL-2000; 2000US-220638P.  
 PR 25-JUL-2000; 2000US-220654P.  
 PR 25-JUL-2000; 2000US-220666P.  
 PR 26-JUL-2000; 2000US-220893P.  
 PR 28-JUL-2000; 2000WO-US20710.  
 PR 23-AUG-2000; 2000WO-US23322.  
 PR 24-AUG-2000; 2000WO-US23328.  
 PR 15-SEP-2000; 2000US-000000P.  
 PR 10-NOV-2000; 2000WO-US30873.  
 PR 28-NOV-2000; 2000US-233646P.  
 PR 01-DEC-2000; 2000WO-US32678.  
 PR 20-DEC-2000; 2000US-0747259.  
 PR 20-DEC-2000; 2000WO-US34956.  
 PR 28-FEB-2001; 2001WO-US06520.  
 PR 10-MAY-2001; 2001US-0854280.  
 PR 25-MAY-2001; 2001WO-US17092.  
 XX  
 XX (GERTH ) GENENTECH INC.  
 PA  
 PI Bakker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;  
 PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Macanabe CK, Wood WI;  
 PI WPI; 2002-112001/22.  
 DR N-PSDB; ABK33575.  
 DR  
 XX One hundred and twenty two nucleic acids encoding PRO polypeptides,  
 PT useful for treating a PRO related disorder and for diagnosing tumours  
 PT such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal  
 PT tumour or liver tumour.  
 XX  
 XX Claim 11, Figure 80; 359pp; English.  
 PS  
 XX The invention relates to one hundred and twenty two nucleic acids  
 CC encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides  
 CC encode human secreted proteins. The PRO nucleic acids, polypeptides,  
 CC agonists and antagonists are useful for treating a PRO related disorder.  
 CC The PRO polypeptides are useful for diagnosing tumours, especially lung  
 CC cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or  
 CC liver tumour. The PRO polypeptides are useful for stimulating the  
 CC proliferation of, or gene expression in, pericyte cells, for stimulating  
 CC the proliferation of, or differentiation of chondrocyte cells, for  
 CC stimulating the release of tumour necrosis factor-alpha from human blood,  
 CC for stimulating or inhibiting the proliferation of normal human dermal  
 CC fibroblast cells. The PRO polypeptide may also be used as molecular  
 CC weight markers and for tissue typing. The PRO nucleic acids have  
 CC applications in molecular biology, including use as hybridisation probes,  
 CC and in chromosome and gene mapping. AAUS3552-AAUS3713 represent human PRO  
 CC protein sequences of the invention.  
 CC  
 XX  
 XX Sequence 667 AA:  
 SQ  
 Query Match 65.5%; Score 1110.5; DB 23; Length 667;  
 Best Local Similarity 76.6%; Pred. No. 1,4e-106;  
 Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;  
 Oy 1 CYSALVLSCLTFLVLMRSVLTNR-----LQSGSGSD-----GQFSNLPFS 42  
 ||||||||||||||||||||||||||||| : | : ||

DB 365 CYSALVLSCLTFLVLMRSVLTNRNLALHRCGAALDLSPLHRS PHSRQAI PCWMSFS 424  
 Oy 43 VPLPLPLAGLLVQOIIIFPLGTTALAPLVMPVHGRVLLFFRSLESSPFKLTALAVY 102  
 DB 425 AYQAFICLGLVQOIIIFPLGTTALAPLVMPVHGRVLLFFRSLESSPFKLTALAVY 484  
 Oy 103 LQNMMAHWFLETHDHPOLTNRRVLYAATFLLEPLNLYGAMXXXC----- 149  
 DB 485 LQNMMAHWFLETHDHPOLTNRRVLYAATFLLEPLNLYGAMVATWRVLLSALYNAIH 544  
 Oy 150 -SPSIAIRP-----TPGYTYRNFPAKIEVSGSPHAMPFCSLLLOQSLPPTMAAPD 203  
 DB 545 GQMDLSLPPRAATLDPPGYTYRNFPAKIEVSGSPHAMPFCSLLLOQSLPPTMAAPD 604  
 Oy 204 SLRGEDEEMQLQTKDSNAGARPAGXRRGARGAYTLNHPLOVFRKTLALGANG 263  
 DB 605 SLRGEDEEMQLQTKDSNAGARPAGXRRGARGAYTLNHPLOVFRKTLALGANG 664  
 Oy 264 AQP 266  
 ||||  
 DB 665 AQP 667  
 ||||  
 RESULT 7  
 ID ABB05048 standard; Protein; 667 AA.  
 XX  
 AC ABB05048;  
 XX  
 DT 25-MAR-2002 (first entry)  
 XX  
 DE Human NOV8c protein SEQ ID NO:32.  
 XX  
 KM Human; NOV8; cytosolic; antidiabetic; anorectic; antibacterial;  
 KM fungicide; vitruicide; protozoanicide; analgesic; antiparkinsonian;  
 KM antisthmatic; hypotensive; osteopathic; antiinflammatory; antilicer;  
 KM neuroprotective; cardiatic; antiallergic; antidepressant; nootropic;  
 KM anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;  
 KM developmental disease; taste and scent detectability disorder; infection;  
 KM Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;  
 KM retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;  
 KM noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;  
 KM hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;  
 KM multiple sclerosis; angina pectoris; myocardial infarction; delirium;  
 KM benign prostatic hypertrophy; manic depression; dementia; dyskinesia;  
 KM severe metal retardation; Huntington's disease; chromosome 15;  
 KM Gilles de la Tourette syndrome.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200190155-A2.  
 PN  
 XX 29-NOV-2001.  
 PD  
 XX 24-MAY-2001; 2001WO-US17073.  
 PF  
 XX 24-MAY-2000; 2000US-206679P.  
 PR 24-MAY-2000; 2000US-206688P.  
 PR 24-MAY-2000; 2000US-206829P.  
 PR 30-MAY-2000; 2000US-207748P.  
 PR 30-MAY-2000; 2000US-207798P.  
 PR 31-MAY-2000; 2000US-208263P.  
 PR 02-JUN-2000; 2000US-208831P.  
 PR 05-JUN-2000; 2000US-209451P.  
 PR 07-JUN-2000; 2000US-210060P.  
 PR 26-JUL-2000; 2000US-219507P.  
 PR 31-JUL-2000; 2000US-221927P.  
 PR 19-JAN-2001; 2001US-263135P.  
 PR 24-JAN-2001; 2001US-263688P.  
 PR 23-MAY-2001; 2001US-0863776.  
 XX



PA (CUBA-) CURAGEN CORP.

XX Stryek KA, Majumder K, Tchernen VT, Mishra V, Padigaru M;  
PI Spaderna SK, Shenoy S, Raetelli L, Li L, Taupier RJ, Garigolli E;  
XX WPI: 2002-106174/14.  
DR N-PSDB: ABA92669.

XX Novel polypeptide, useful for treating pain, cancer, urinary retention,  
PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,  
PT asthma, ulcer, allergy and Huntington's disease, comprises isolated  
PT polypeptide NOVX -

PS Claim 1, Page 106, 266pp; English.

XX The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX  
CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,  
CC fungicide, virucide, protozoacide, analgesic, antiparkinsonian,  
CC antiaesthetic, hypotensive, osteopathic, antiinflammatory, antitumor,  
CC neuroprotective, cardiotonic, antiallergic, antidepressant, nootropic,  
CC anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)  
CC can be used in gene therapy and vaccine production. (I) and (II) can be  
CC used for treating or preventing a NOVX-associated disorder such as  
CC cardiomyopathy, atherosclerosis and diabetes in a human, where the  
CC disorder is related to cell signal processing and metabolic path way  
CC modulation, in a subject, preferably human. (I) and (II) can be used for  
CC diagnosing, preventing or treating developmental diseases, immune  
CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,  
CC signal transduction pathway disorders, retinal diseases including those  
CC involving photoreception, cell growth rate disorders, feeding disorders,  
CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,  
CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,  
CC Albright Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial  
CC infection, allergies, benign prostatic hypertrophy, manic depression,  
CC delirium, dementia, severe mental retardation and dyskinesias, such as  
CC Huntington's disease or Gilles de la Tourette syndrome. The present  
CC sequence represents the human NOVX protein from the present invention.  
CC Human NOVX2 is isolated to chromosome 15.  
XX

XX Sequence 667 AA;

Query Match 65.5%; Score 1110.5; DB 23; Length 667;  
Best Local Similarity 76.6%; Pred. NO. 1.4e-106;  
Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYISALVLSCLTFLVLRSLVYTR-----LGSQSGSD-----GQFSWNLFS 42  
DB 365 CYISALVLSCLTFLVLRSLVYTRNLRAHRAALDLSPIHSPHPSQALFCWMSFS 424  
QY 43 VPLPLPLAGLVVQOIIFFLGTALAEVLPVYHGRLLPFRGLSSNPFMTALAVI 102  
DB 425 AVGTAFICGLLVQOIIFFLGTALAEVLPVYHGRLLPFRGLSSNPFMTALAVI 484  
QY 103 LQNMAMHVFLETHDGHPLTNRVLYATFLLPPLNVLYGAMKXXC----- 149  
DB 485 LQNMAMHVFLETHDGHPLTNRVLYATFLLPPLNVLYGAMVATRWVLSALVNAIHL 544  
QY 150 -SSSLAIRP-----TPGYTYRNLKLEVGOSHPAMTAFCSLLLOQSLPPTMAFPD 203  
DB 545 GQMDLSLPLPRAATLDPGYTYRNLKLEVGOSHPAMTAFCSLLLOQSLPPTMAFPD 604  
QY 204 SLRPGEDGQMLQTKQSMKAGARPGAXRGRARWGLATYLLHPTLQVFRKALGANG 263  
DB 605 SLRPGEDGQMLQTKQSMKAGARPGASRGRARWGLATYLLHPTLQVFRKALGANG 664  
QY 264 AQP 266  
DB 665 AQP 667

RESULT 8  
AAU91278

ID AAU91278 standard; Protein, 689 AA.

XX AAU91278;

AC 18-JUN-2002 (first entry)

DE Human NOV2b protein.

XX Human: NOVX; gene therapy; cardiomyopathy; atherosclerosis;  
XX diabetes; cell signal processing; metabolic pathway modulation;  
XX inflammation; autoimmune disorder; scleroderma; transplantation;  
XX allergy; systemic lupus erythematosus; haemophilia; Alzheimer's disease;  
XX graft versus host disease; Leisch-Nyhan syndrome; periodontitis;  
XX pancreatitis; musculoskeletal disorder; Parkinson's disease;  
XX Huntington's disease; behavioural disorder; pain; obesity; wound healing;  
XX neurodegenerative disorder; neuropsychiatric disorder; hypertension;  
XX growth disorder; reproductive disorder; lung disease.

OS Homo sapiens.

PN MO200216600-A2.

XX 28-FEB-2002.

PF 27-AUG-2001; 2001WO-US26518.

XX 25-AUG-2000; 2000US-227800P.

PR 25-AUG-2000; 2000US-228205P.

PR 25-AUG-2000; 2000US-228324P.

PR 30-AUG-2000; 2000US-228997P.

PR 30-AUG-2000; 2000US-229185P.

PR 01-SEP-2000; 2000US-229780P.

PR 01-SEP-2000; 2000US-229850P.

PR 22-JUN-2001; 2001US-263337P.

PR 31-JUN-2001; 2001US-265518P.

PR 15-MAR-2001; 2001US-276451P.

PR 27-MAR-2001; 2001US-279196P.

PR 24-AUG-2001; 2001US-0393398.

XX (CUBA-) CURAGEN CORP.

XX Gerlach V, McDougall JR, Smithson G, Stone DJ, Ellerman K;

PI Stryek KA, Zernhsen BD, Raetelli L, Verney CM, Patutajan M;

PI Tchernen VT, Padigaru M, Taupier RJ;

XX WPI: 2002-292064/33.

DR N-PSDB: ABA92669.

PT New isolated cytoplasmic, nuclear, membrane bound and secreted

PT polypeptides, termed NOVX, useful for treating inflammation, autoimmune

PT disorders, haemophilia, Leisch-Nyhan syndrome, pancreatitis,

PT musculoskeletal disorders -

XX Claim 1, Page 20, 245pp; English.

XX The invention relates to an isolated cytoplasmic, nuclear, membrane bound

CC or secreted polypeptide, designated NOVX (actually NOV1, 2a, 2b, 3a, 3b,

CC 4, 5a, 5b, 5c, 5d, 5e, 5f, 5g, 5h, 5i, 6, 7 and 8), a variant of NOVX, a

CC mature form, or a variant of the mature form of NOVX. Also included

CC are a polynucleotide encoding NOVX (or its complement), a vector

CC comprising the polynucleotide, a cell comprising the vector, an

CC anti-NOVX antibody, determining the presence of NOVX in a sample

CC using the antibody, determining the presence of NOVX polynucleotide in

CC a sample using a probe which binds to NOVX polynucleotide, identifying a

CC an agent which binds to NOVX (including modulators of NOVX). NOVX, the

CC polynucleotide and the antibody are useful for diagnosing, treating

CC or preventing a NOVX-associated disorder selected from cardiomyopathy,

CC atherosclerosis, diabetes, a disorder related to cell signal processing

CC and metabolic pathway modulation, inflammation, autoimmune disorders,

CC scleroderma, transplantation, allergies, systemic lupus erythematosus,

CC haemophilia, graft versus host disease, Alzheimer's disease, stroke,

CC Leisch-Nyhan syndrome, periodontitis, pancreatitis, musculoskeletal

CC disorders; Parkinson's disease, Huntington's disease, behavioural  
 CC disorders; pain, neurodegenerative and neuropsychiatric disorders,  
 CC hypertension, wound healing, obesity, growth and reproductive  
 CC disorders, lung diseases and many other diseases and antibodies listed in  
 CC the specification. NOVX, the polynucleotide and the antibody are useful  
 CC in screening assays, detection assays (e.g., chromosomal mapping, tissue  
 CC typing, forensic biology), predictive medicine (e.g., diagnostic assays,  
 CC prognostic assays, monitoring clinical trials and pharmacogenomics), and  
 CC in methods of treatment (e.g., therapeutic and prophylactic). NOVX is  
 CC useful as immunogen to produce antibodies immunospecific for NOVX, as  
 CC vaccines to screen for potential agonist and antagonist compounds, and  
 CC as bait protein in a two-hybrid or three-hybrid assay. The polynucleotide  
 CC is useful in gene therapy, to express NOVX, to detect NOVX mRNA  
 CC or a genetic lesion in a NOVX gene, and to modulate NOVX activity. The  
 CC vector is useful for producing non-human transgenic animals. The antibody  
 CC is useful for isolating, and purifying NOVX and to monitor protein levels  
 CC in tissue as part of a clinical testing procedure. The present sequence  
 CC represents a NOVX protein.

XX Sequence 689 AA;

Query Match 65.5%; Score 1110.5; DB 23; Length 689;  
 Best Local Similarity 76.6%; Pred. No. 1.5e-106;  
 Matches 232; Conservative 4; Mismatches 30; Indels 37; Gaps 4;

QY 1 CYSALVLSCLTFVLMPSLTHR-----LGSQSGSP-----GQFSWNTFS 42  
 DB 387 CYSALVLSCLTFVLMPSLTHRNRALHRAALGALSLHRSPPSRQALFCWMSFS 446  
 QY 43 VPLPLPPLAGLLVQOIIFFLGTALAPLVLPVHGRNLLPFRSLSSNPFWTLALAVI 102  
 DB 447 AVGTATCICGLLVQOIIFFLGTALAPLVLPVHGRNLLPFRSLSSNPFWTLALAVI 506  
 QY 103 LQNMAMHWVLETHDHPOLTRRVLYATFLPPLVNLGAMKXXC----- 149  
 DB 507 LQNMAMHWVLETHDHPOLTRRVLYATFLPPLVNLGAMKXXC----- 149  
 QY 150 -SPSIAIRHP-----TPGYTYRNFLEKIEVSQSPHMTAFCSLLLOQSLPRTMAPOD 203  
 DB 567 GQMDLPLPRATLDPGYTYRNFLEKIEVSQSPHMTAFCSLLLOQSLPRTMAPOD 626  
 QY 204 SLRPGEDSGMOLLQTKSMAGKARPGAXRGRARNGLYATLHNPLOVRKTKALLGANG 263  
 DB 627 SLRPGEDSGMOLLQTKSMAGKARPGAXRGRARNGLYATLHNPLOVRKTKALLGANG 686  
 QY 264 AQP 266  
 DB 687 AQP 689

RESULT 9  
 AAU04558  
 ID AAU04558 standard; Protein; 658 AA.

XX AAU04558;

XX 26-SBP-2001 (first entry)

XX Human Str66 homologue, PRO19578.

XX Human; Str66, PRO19578, 15q23; tumour; breast cancer;  
 KW head tumour; leukaemia; lymphoid malignancy; inflammatory disorder;  
 XX immunogenic disorder; antigen; antibody.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Modified-site 8..12 /note= "Asn is N-glycosylated"

XX Region 50..56 /label= N\_myristoylation\_site

XX Domain 54..71 /label= Transmembrane\_domain

FT Domain 93..111  
 FT /label= Transmembrane\_domain  
 FT 123..131 /note= "ATP/GTP-binding site motif A (P-loop)"  
 FT 140..157  
 FT /label= Transmembrane\_domain  
 FT 167..173  
 FT /label= N\_myristoylation\_site  
 FT 197..214  
 FT /label= Transmembrane\_domain  
 FT 232..238  
 FT /label= N\_myristoylation\_site  
 FT 291..312  
 FT /label= Transmembrane\_domain  
 FT 308..314  
 FT /label= N\_myristoylation\_site  
 FT 332..338  
 FT /label= N\_myristoylation\_site  
 FT 355..366  
 FT /note= "Prokaryotic membrane lipoprotein attachment site"  
 FT 356..371  
 FT /label= Transmembrane\_domain  
 FT 425..444  
 FT /label= Transmembrane\_domain  
 FT 464..481  
 FT /label= Transmembrane\_domain  
 FT 505..522  
 FT /label= Transmembrane\_domain  
 FT 516..522  
 FT /label= N\_myristoylation\_site  
 FT 618..624  
 FT /label= N\_myristoylation\_site  
 FT 622..628  
 FT /label= N\_myristoylation\_site  
 FT 631..637  
 FT /label= N\_myristoylation\_site  
 FT 652..658  
 FT /label= N\_myristoylation\_site

PN MO200151635-A2.

PD 19-JUL-2001.

XX 11-JAN-2001; 2001WO-US00847.

XX 13-JAN-2000; 2000US-0175849.

PR 14-APR-2000; 2000US-0197089.

PR 29-AUG-2000; 2000US-0228914.

XX (GETH ) GENENTECH INC.

XX Pennica D, Smith V, Wood WI,

XX MPI; 2001-442146/47.

XX N-PSDB; AAS08632.

XX Human nucleic acid encoding a PRO10282 or PRO19578 polypeptide (the PRO

XX polypeptides are referred to as Str66 polypeptides), useful in

XX molecular biology, including use as hybridisation probes, and in

XX chromosome and gene mapping.

XX Claim 24; Fig 7; 1599P; English.

CC The sequence is human PRO19578 (a variant possibly due to  
 CC alternative splicing of the Str66 mRNA), a homologue of murine Str66, a  
 CC retinole acid responsive protein thought to play an important role in  
 CC early dorsoventral limb patterning during development and later in the  
 CC control of endochondral ossification. The gene for the Str66 homologue  
 CC is located on chromosome 15q23. The Str66 polypeptides, agonists,  
 CC antagonists or anti-Str66 antibodies are useful for preparation of a  
 CC medicament useful in the treatment of a condition which is responsive to  
 CC the Str66 polypeptides, agonists, antagonists or anti-Str66 antibodies.

The Strept polypeptides may also be employed as molecular weight markers for protein electrophoresis. The Stre nucleic acids have applications in molecular biology, including use as hybridisation probes, and in chromosome and gene mapping. The antibodies and other anti-tumour compounds may be used to treat various conditions, including those characterised by overexpression and/or activation of amplified genes. Exemplary conditions or disorders include benign or malignant tumours (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulvar, thyroid, hepatic carcinomas, sarcomas, glioblastomas, and various head and neck tumours, leukemias and lymphoid malignancies), neuronal, glial, astrocytal, hypothalamic and other glandular, macrophagal, epithelial, stromal and blastococic disorders, inflammatory, angiogenic and immunogenic disorders.

**SQ** Sequence 658 AA;

Query Match	65.3%;	Score 1106.5;	DB 22;	Length 658;
Best Local Similarity	76.2%;	Pred. No. 3.6e-106;		
Matches 231; Conservative	5;	Mismatches 30;	Indels 37;	Gaps 4;

QY	1	CYISALVISCLETPFLVMRSIVYTR-----IGSGGSD-----GQSNMLES	42
Db	356	CYISALVISCLETPFLVMRSIVYTRSLVYTRMTLRALHRCALADISPLHRSPHRSRONITCOMSFS	415
QY	43	VPLRLPPLAGLLVQOIIFFLGTTALFLVMPVLHGNNLLFRSDSSNPMULTALAVI	102
Db	416	AYOTAFICIGLLVQOIIFFLGTTALFLVMPVLHGNNLLFRSDSSNPMULTALAVI	475
QY	103	LQMAAAHWFLFLETHGHQPOLNNRRLVYAAFPLLPLNVLVGMXXXC-----	149
Db	476	LQMAAAHWFLFLETHGHQPOLNNRRLVYAAFTPLPLNVLVGMIVATRWVLSALYMAIHL	535
QY	150	-SPSIAIRHP-----TPGYIYTRNPLKIEVSQSHPMATPACSLILQAQSLPRTMAAPQD	203
Db	536	GQMDLSLTPPRATLDPGYIYTRNPLKIEVSQSHPMATPACSLILQAQSLPRTMAAPQD	595
QY	204	SLRPGEDSGMOLLQTKOSMAKGARPGAXXGGRARWGLATLTLHNPTLOVPRKTLGANG	263
Db	596	SLRPGEDSGMOLLQTKOSMAKGARPGAXXGGRARWGLATLTLHNPTLOVPRKTLGANG	655
QY	264	AQP 266	
Db	656	AQP 658	

RESULT 10  
AAU78576  
ID AAU78576 standard; Protein; 658 AA.

AC	AAU79576;	
XX		
DT	18-JUN-2002	(first entry)
XX		
DE	Human Stra6 protein variant (PRO19578) .	
XX		
KM	Human; cancer; Stra6; PRO19578; cytosolic; stromal disorder;	
KM	tumour; retinoid; Mnt; tumour antigen; leukaemia; lymphoid malignancy	
KM	neural disorder; glial disorder; astrocyta disorder;	
KM	hypothalamic disorder; glandular disorder; macrophagal disorder;	
KM	epithelial disorder; blasococelic disorder; chromosome 15q23.	
XX		
OS	Homo sapiens.	
XX		
Key	Location/Qualifiers	
FT	8..12	
FT	/note= "N Glycosylation site"	
FT	50..56	
FT	/note= "N-myristoylation site"	
FT	54..71	
FT	/note= "Transmembrane domain"	
FT	93..111	
FT	/note= "Transmembrane domain"	
Domain		

FT	Region	123..131
FT	/note="ATP/GTP binding site motif A"	
FT	Domain	140..157
FT	/note="Transmembrane domain"	
FT	Region	167..173
FT	/note="N-myristoylation site"	
FT	Domain	197..214
FT	/note="Transmembrane domain"	
FT	Region	232..238
FT	/note="N-myristoylation site"	
FT	Domain	291..312
FT	/note="Transmembrane domain"	
FT	Region	308..314
FT	/note="N-myristoylation site"	
FT	Region	332..338
FT	/note="N-myristoylation site"	
FT	Domain	356..371
FT	/note="Transmembrane domain"	
FT	Region	355..366
FT	/note="lipoprotein lipid attachment site"	
FT	Domain	425..444
FT	/note="Transmembrane domain"	
FT	Domain	464..481
FT	/note="Transmembrane domain"	
FT	Region	516..522
FT	/note="N-myristoylation site"	
FT	Domain	505..522
FT	/note="Transmembrane domain"	
FT	Region	618..628
FT	/note="N-myristoylation site"	
FT	Region	622..628
FT	/note="N-myristoylation site"	
FT	Region	631..637
FT	/note="N-myristoylation site"	
FT	Region	652..658
FT	/note="N-myristoylation site"	
PN	WO200218608-A2.	
XX		
XX		
XX	07-MAR-2002.	
PD		
XX		
PR	10-JUL-2001; 2001WO-US21635.	
XX		
PR	29-AUG-2000; 2000US-228914P.	
XX		
PR	11-JAN-2001; 2001US-0759056.	
XX		
PR	10-JUL-2001; 2001US-0901812.	
XX		
PA	(GETH ) GENENTECH INC.	
XX		
PI	Tice D, Pennica D, Polakis P, Szeto W;	
XX		
XX	WPI; 2002-292202/33.	
DR	N-PSDB; ABK47280.	
XX		
XX	Selectively enhancing the expression of a protein in a tumour cell	
PT	characterised by aberrant Wnt signalling, useful for enhancing the	
PT	efficacy of treatment of cancers, comprises treating the tumour cell	
PT	with a retinoid -	
XX		
XX		
PS	Example 1; Fig 7; 14pp; English.	
XX		
CC	This invention relates to a novel method for selective enhancement of	
CC	the expression of a protein in a tumour cell characterised by aberrant	
CC	Wnt signalling by treating the tumour cell with a retinoid. The method	
CC	is useful for enhancing the efficacy of treatment of cancers	
CC	characterised by aberrant Wnt signalling, and for identifying tumour	
CC	antigens the expression of which is selectively enhanced by retinoid	
CC	treatment. The method is also useful in the early detection of	
CC	tumours for early intervention. The tumour antigens are useful as	
CC	targets for cancer therapy, since selective enhancement of their	
CC	expression by retinoid treatment relative to normal cells improves the	
CC	efficacy and therapeutic index of cancer therapeutics directed against	
CC	these antigens. Antibodies binding to the tumour antigens may be	

CC administered for the treatment of various disorders or conditions,  
 CC including those characterized by overexpression and/or activation of the  
 CC tumour antigens, where such conditions or disorders include benign or  
 CC malignant tumours, leukemias and lymphoid malignancies; neuronal, glial,  
 CC astrocytic, hypothalamic and other glandular, macrophagal, epithelial,  
 CC stromal and blastocoele disorders. The present sequence represents  
 CC the human Str6 protein variant (PRO19578) which is actively upregulated  
 CC by retinoid using the method of the invention. The gene encoding the  
 CC human Str6 protein is located on human chromosome 15q23.

XX Sequence 658 AA;

Query Match 65.3%; Score 1106.5; DB 23; Length 658;

Best Local Similarity 76.2%; Pred. No. 3.6e-106; Mismatches 231; Conservative 5; Indels 37; Gaps 4;

QY 1 CYSALVLSGLTFVLVMSLVYTHR-----IGSGGSGD-----GQFSNNLFS 42

DB 356 CYSALVLSGLTFVLVMSLVYTHRNRALHRAALDSLPHRSPPSOALFCMWSFS 415

QY 43 VPLPPLAGLVQOIIFFLGTALAFVLMFVHGRNLLPFRSLSSWPFMTLALAVI 102

DB 416 AYOTAFICGLVQOIIFFLGTALAFVLMFVHGRNLLPFRSLSSWPFMTLALAVI 475

QY 103 LQNMAMHWVFLTHDHPQLTRRVLYAATFLPPLNVGAMXXXC----- 149

DB 476 LQNMAMHWVFLTHDHPQLTRRVLYAATFLPPLNVGAMXXXC----- 535

QY 150 -SPSIAIRHP-----TPGYTYRNPFLKIEVSGSHPMATFCSLLAQSLPPTMAPOD 203

DB 536 GQNDLSLPRATDPGYTYRNPFLKIEVSGSHPMATFCSLLAQSLPPTMAPOD 595

QY 204 SLRPGDEDMQLQTKDSMARGARPGARPGVGLAYTLNPTLOVRKTLALGANG 263

DB 596 SLRPGDEDMQLQTKDSMARGARPGARPGVGLAYTLNPTLOVRKTLALGANG 655

QY 264 AQP 266

DB 656 AQP 658

RESULT 11

ID ABB05046 standard; Protein; 653 AA.

XX ABB05046;

DT 25-MAR-2002 (first entry)

XX Human NOV8a protein SEQ ID NO:28.

XX Human; NOVX; cytostatic; antidiabetic; anorectic; antibacterial;  
 XX fungicide; virocidic; protozoacide; analgesic; antiparkinsonian;  
 XX antisthmatic; hypotensive; osteopathic; antiinflammatory; antitumor;  
 XX neuroprotective; cardiant; antiallergic; antidepressant; nootropic;  
 XX anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;  
 XX developmental disease; taste and scent detectability disorder; infection;  
 XX Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;  
 XX retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;  
 XX noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;  
 XX hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;  
 XX multiple sclerosis; angina pectoris; myocardial infarction; delirium;  
 XX benign prostatic hypertrophy; manic depression; dementia; dyskinesia;  
 XX severe metal retardation; Huntington's disease; chromosome 15;  
 XX Gilles de la Tourette syndrome.

XX Homo sapiens.

XX WO200190155-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US17073.

XX 24-MAY-2001; 2000US-206679P.  
 PR 24-MAY-2001; 2000US-206688P.  
 PR 24-MAY-2001; 2000US-206829P.  
 PR 30-MAY-2001; 2000US-207748P.  
 PR 30-MAY-2001; 2000US-207798P.  
 PR 31-MAY-2001; 2000US-208263P.  
 PR 02-JUN-2001; 2000US-208831P.  
 PR 05-JUN-2001; 2000US-209453P.  
 PR 07-JUN-2001; 2000US-210060P.  
 PR 20-JUL-2001; 2000US-219507P.  
 PR 26-JUL-2001; 2000US-221337P.  
 PR 31-JUL-2001; 2000US-221927P.  
 PR 19-JAN-2001; 2001US-263135P.  
 PR 24-JAN-2001; 2001US-263688P.  
 PR 24-JAN-2001; 2001US-263694P.  
 PR 23-MAY-2001; 2001US-0863776.  
 PR (CURA-) CURAGEN CORP.  
 PR Spletz KA, Majumder K, Tchernev VT, Mishra V, Padigaru M;  
 PI Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RJ, Gangolli E;  
 XX WPI: 2002-106174/14.  
 DR N-FSDB; ABA92663.

PT Novel polypeptide, useful for treating pain, cancer, urinary retention,  
 PT osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,  
 PT asthma, ulcer, allergy and Huntington's disease, comprises isolated  
 PT polypeptide NOVX -  
 XX  
 PS Claim 1; Page 98; 266pp; English.

XX The present invention describes human NOVX proteins (NOV1 to NOV9). NOVX  
 CC proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,  
 CC fungicide, virocidic, protozoacide, analgesic, antiparkinsonian,  
 CC antisthmatic, hypotensive, osteopathic, antiinflammatory, antitumor,  
 CC neuroprotective, cardiant, antiallergic, antidepressant, nootropic, (II)  
 CC anticonvulsant and neuroleptic activities. (I) and polymeric forms (II)  
 CC can be used in gene therapy and vaccine production. (I) and (II) can be  
 CC used for treating or preventing a NOVX-associated disorder such as  
 CC cardiomyopathy, atherosclerosis and diabetes in a human, where the  
 CC disorder is related to cell signal processing and metabolic path way  
 CC modulation, in a subject, preferably human. (I) and (II) can be used for  
 CC diagnosing, preventing or treating developmental diseases, immune  
 CC diseases, taste and scent detectability disorder, Burkitt's lymphoma,  
 CC signal transduction pathway disorders, retinal diseases including those  
 CC involving photoreception, cell growth rate disorders, feeding disorders,  
 CC noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,  
 CC Albright hereditary osteodystrophy, angina pectoris, ulcers, myocardial  
 CC infarction, allergy, benign prostatic hypertrophy, manic depression,  
 CC delirium, dementia, severe metal retardation and dyskinesias, such as  
 CC Huntington's disease or Gilles de la Tourette syndrome. The present  
 CC sequence represents the human NOV8a protein from the present invention.  
 CC Human NOV2a is isolated to chromosome 15.

XX Sequence 653 AA;

Query Match 62.6%; Score 1060.5; DB 23; Length 653;

Best Local Similarity 74.6%; Pred. No. 2.2e-101; Mismatches 226; Conservative 4; Mismatches 30; Indels 43; Gaps 6;

QY 1 CYSALVLSGLTFVLVMSLVYTHR-----IGSGGSGD-----GQFSNNLFS 42

DB 357 CYSALVLSGLTFVLVMSLVYTHRNRALHRAALDSLPHRSPPSOALFCMWSFS 416

QY 43 VPLPPLAGLVQOIIFFLGTALAFVLMFVHGRNLLPFRSLSSWPFMTLALAVI 102

DB 417 AYOTAFICGLVQOIIFFLGTALAFVLMFVHGRNLLPFRSLSSWPFMTLALAVI 470

QY 103 LQNMAMHWVFLTHDHPQLTRRVLYAATFLPPLNVGAMXXXC----- 149

```

Db 471 LQNMAAHWVPLETHDGHQPOLNRRVLYAATFLPPLVNLVGAVATWVRVLSALVNAIHL 530
Qy 150 -SPSIAIRHP-----TRGYTYRRNFKLEIVSQSHPMATACSLLLQAOSSLPRMTMAAPD 203
Db 531 GQMDLSLPRPAATLDPGYTYRRNFKLEIVSQSHPMATACSLLLQAOSSLPRMTMAAPD 590
Qy 204 SLRPGEDGMOQLQTKDSMAKGPAGXRGARWGIAVYTLNPTLQVFRKTALLGANG 263
Db 591 SLRPGEDGMOQLQTKDSMAKGPAGXRGARWGIAVYTLNPTLQVFRKTALLGANG 650
Qy 264 AQP 266
Db 651 AQP 653

RESULT 12
ABR05047
ID ABR05047 standard; Protein: 662 AA.
AC ABR05047;
DT 25-MAR-2002 (first entry)
DE Human NOV8b protein SEQ ID NO:30.
XX
XX Human; NOVX; cytostatic; antidiabetic; anorectic; antibacterial;
XX fungicide; virucide; protozoacide; analgesic; antiparkinsonian;
XX antipneumonic; hypotensive; osteopathic; antiinflammatory; antitumor;
XX neuroprotective; cardiant; antiallergic; antidepressant; nootropic;
XX anticonvulsant; neuroleptic; gene therapy; vaccine; immune disease;
XX developmental disease; taste and scent detectability disorder; infection;
XX Burkitt's lymphoma; signal transduction pathway disorder; pain; cancer;
XX retinal disease; feeding disorder; asthma; Parkinson's disease; ulcer;
XX noninsulin-dependent diabetes mellitus; acute heart failure; hypotension;
XX hypertension; urinary retention; osteoporosis; Crohn's disease; allergy;
XX multiple sclerosis; angina pectoris; myocardial infarction; delirium;
XX benign prostatic hypertrophy; manic depression; dementia; dyskinesia;
XX severe mental retardation; Huntington's disease;
XX Gilles de la Tourette syndrome.
XX
XX Homo sapiens.
XX
XX MO200190155-A2.
XX
XX 29-NOV-2001.
XX
XX 24-MAY-2001; 2001WO-US17073.
XX
XX 24-MAY-2000; 2000US-206679P.
XX 24-MAY-2000; 2000US-206688P.
XX 24-MAY-2000; 2000US-206829P.
XX 30-MAY-2000; 2000US-207748P.
XX 30-MAY-2000; 2000US-207798P.
XX 31-MAY-2000; 2000US-208263P.
XX 02-JUN-2000; 2000US-208831P.
XX 05-JUN-2000; 2000US-209451P.
XX 07-JUN-2000; 2000US-210060P.
XX 20-JUL-2000; 2000US-219507P.
XX 26-JUL-2000; 2000US-221337P.
XX 31-JUL-2000; 2000US-221927P.
XX 19-JAN-2001; 2001US-263135P.
XX 24-JAN-2001; 2001US-263688P.
XX 24-JAN-2001; 2001US-263694P.
XX 23-MAY-2001; 2001US-0863776.
XX
XX (CURA-) CURAGEN CORP.
XX
XX SPYtek KA, Majumder K, Tchernav VT, Mishra V, Padigaru M,
XX Spaderna SK, Shenoy S, Rastelli L, Li L, Taupier RO, Gangolli E;
XX WPI; 2002-106174/14.
XX
XX N-PSDB; ABA92666.
XX

```

```

XX Novel polypeptide, useful for treating pain, cancer, urinary retention,
XX osteoporosis, Crohn's disease, diabetes, acute heart failure, dementia,
XX asthma, ulcer, allergy and Huntington's disease, comprises isolated
XX polypeptide NOVX -
XX
XX Claim 1; Page 103; 266pp; English.
XX
XX The present invention describes human NOVX protein (NOV1 to NOV9). NOVX
XX proteins (I) have cytostatic, antidiabetic, anorectic, antibacterial,
XX fungicide, virucide, protozoacide, analgesic, antiparkinsonian,
XX antipneumonic, hypotensive, osteopathic, antiinflammatory, antitumor,
XX neuroprotective, cardiant, antiallergic, antidepressant, nootropic,
XX anticonvulsant and neuroleptic activities. (I) and polynucleotides (II)
XX can be used in gene therapy and vaccine production. (I) and (II) can be
XX used for treating or preventing a NOVX-associated disorder such as
XX cardiomyopathy, atherosclerosis and diabetes in a human, where the
XX disorder is related to cell signal processing and metabolic path way
XX modulation, in a subject, preferably human. (I) and (II) can be used for
XX diagnosing, preventing or treating developmental diseases, immune
XX diseases, taste and scent detectability disorder, Burkitt's lymphoma,
XX signal transduction pathway disorders, retinal diseases including those
XX involving photoreception, cell growth rate disorders, feeding disorders,
XX noninsulin-dependent diabetes mellitus, infections, pain, cancer, asthma,
XX Parkinson's disease, acute heart failure, hypotension, hypertension,
XX urinary retention, osteoporosis, Crohn's disease, multiple sclerosis,
XX Alibright Hereditary Osteodystrophy, angina pectoris, ulcers, myocardial
XX infarction, allergies, benign prostatic hypertrophy, manic depression,
XX delirium, dementia, severe mental retardation and dyskinesias, such as
XX Huntington's disease or Gilles de la Tourette syndrome. The present
XX sequence represents the human NOV8b protein from the present invention.
XX
XX SQ Sequence 662 AA;
XX
XX Query Match 62.6%; Score 1060.5; DB 23; Length 662;
XX Best Local Similarity 74.6%; Pred. No. 2,2e-101;
XX Matches 226; Conservative 4; Mismatches 30; Indels 43; Gaps 6;
XX
XX Qy 1 CYISALVNSCLTTPVLNRSLVYTHR-----LGSQSGSD-----GQFSWNLFS 42
XX Db 366 CYISALVNSCLTTPVLNRSLVYTHRNLRALHGAALDLSPLHRSPHPRQAI FCMSPS 425
XX
XX Qy 43 VPLPLPLAGLIVQOIIFFLGCTTAFLVLMPLYHGRNLIFFRSLSMSPFMTALAVI 102
XX Db 426 AYQTAFLICGLIVQOIIFFLGCTTAFLVLMPLYHGRNLIFFRSLSMSPFMTALAVI 102
XX
XX Qy 103 LQNMAAHWVPLETHDGHQPOLNRRVLYAATFLPPLVNLVGAVATWVRVLSALVNAIHL 149
XX Db 480 LQNMAAHWVPLETHDGHQPOLNRRVLYAATFLPPLVNLVGAVATWVRVLSALVNAIHL 539
XX
XX Qy 150 -SPSIAIRHP-----TRGYTYRRNFKLEIVSQSHPMATACSLLLQAOSSLPRMTMAAPD 203
XX Db 540 GQMDLSLPRPAATLDPGYTYRRNFKLEIVSQSHPMATACSLLLQAOSSLPRMTMAAPD 599
XX
XX Qy 204 SLRPGEDGMOQLQTKDSMAKGPAGXRGARWGIAVYTLNPTLQVFRKTALLGANG 263
XX Db 600 SLRPGEDGMOQLQTKDSMAKGPAGXRGARWGIAVYTLNPTLQVFRKTALLGANG 659
XX
XX Qy 264 AQP 266
XX Db 660 AQP 662

RESULT 13
AAM23581
ID AAM23581 standard; Protein: 247 AA.
AC AAM23581;
DT 12-OCT-2001 (first entry)
DE Murine EST encoded protein SEQ ID NO: 1106.
XX
XX
XX

```

Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse; toad; monkey; dog; sea urchin; expressed sequence tag; EST; diagnostics; forensic test; gene mapping; genetic disorder; biodiversity; gene therapy; nutrition.

Mus musculus.

WO200154477-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US02687.

25-JAN-2000; 2000US-0491404.

17-JUL-2000; 2000US-0617746.

03-AUG-2000; 2000US-0631451.

15-SEP-2000; 2000US-0663870.

(HYSE-) HYSEQ INC.

Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V; Cao Y, Drmanac RA, Zhang J, Werhman T;

WPI: 2001-476164/51.

N-PSDB; AAC98240.

Isolated polypeptide for treatment of diseases, diagnostics, raising antibodies and research use -

Claim 20; Page 825-826; 1275pp; English.

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and toad. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a protein of the invention.

Sequence 247 AA:

Query Match 59.1%; Score 1002.5; DB 22; Length 247; Best Local Similarity 85.5%; Pred. No. 6.6e-96; Matches 200; Conservative 3; Mismatches 12; Indels 19; Gaps 2;

52 GLVYQIIFFLGTTLALAFVLPVLPVHGRNLLFFRSLESSWPFMTLALAVILQNMAMHW 111  
14 GLVYQIIFFLGTTLALAFVLPVLPVHGRNLLFFRSLESSWPFMTLALAVILQNMAMHW 73

112 PLETHDGHPOLTNRRVLYAATFLLPPLNVLVGAMXXXC-----SPSIAIRH 157  
74 PLETHDGHPOLTNRRVLYAATFLLPPLNVLVGAMVATRVVLSALYNAIHVGQNDLSLP 133

158 P-----TPGYTYRNFLEKIEVSQSHPMATFCSLLLOQSLPRTMAPODSLRPEEDE 212  
134 PRAATIDRGVYTRNFKIEVSQSHPMATFCSLLLOQSLPRTMAPODSLRPEEDE 193

213 GMQLLQTKDSMAGARPGAXRGARWGLAYTILNPTQVRRKTLGANGAQC 266  
194 GMQLLQTKDSMAGARPGASRGARWGLAYTILNPTQVRRKTLGANGAQC 247

RESULT 14  
AAB53256  
ID AAB53256 standard; Protein; 178 AA.  
XX  
AC AAB53256;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human colon cancer antigen protein sequence SEQ ID NO:796.  
XX

Human; colon cancer; colon cancer antigen; diagnosis; detection; identification; cytotoxic; cardioactive; neuroprotective; vulnery; immunomodulatory; muscular; gynaecological; gastrointestinal; nephrotropic; anti-infective; antibacterial; gene therapy; wound; neural disorder; immune system disorder; muscular disorder; reproductive disorder; gastrointestinal disorder; renal disorder; infectious disease; cardiovascular disorder.

Homo sapiens.

WO200055351-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05883.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2000-587534/55.

N-PSDB; AAC98013.

Colon cancer associated gene sequences, referred to as colon cancer antigens, useful for the treatment, prevention, and diagnosis of colon disorders such as colon cancer -

Claim 11; Page 1352-1353; 2104pp; English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in AAB5334 to AAB54006. The human colon cancer antigens can have cytotoxic, cardioactive, muscular; neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and vulnery, nephrotropic, anti-infective and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The polynucleotides may be used in diagnostics and research, such as for chromosome identification, and as hybridization probes. The proteins may also be used to prevent diseases such as neural disorders, immune system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, wounds, renal disorders, infectious diseases, and cardiovascular disorders. AAC9764 to AAC9872 and AAB54007 represent sequences used in the exemplification of the present invention.

Sequence 178 AA:

Query Match 28.6%; Score 485; DB 21; Length 178; Best Local Similarity 88.2%; Pred. No. 3.1e-42; Matches 97; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

36 FSNWLVPLPPLAGLIVQIIFFLGTTLALAFVLPVLPVHGRNLLFFRSLESSWPFML 95  
27 FSNWLVPLPPLAGLIVQIIFFLGTTLALAFVLPVLPVHGRNLLFFRSLESSWPFML 86

96 TLALAVILQNMAMHWPLETHDGHPOLTNRRVLYAATFLLPPLNVLVGAM 145  
87 TLALAVILQNMAMHWPLETHDGHPOLTNRRVLYAATFLLPPLNVLVGAM 136

RESULT 15  
AAG74057  
ID AAG74057 standard; Protein; 178 AA.  
XX  
AC AAG74057;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen protein SEQ ID NO:4821.  
XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW colorectal carcinoma.  
 OS Homo sapiens.  
 XX MO200122920-A2.  
 XX  
 PD 05-APR-2001.  
 XX  
 PF 28-SEP-2000; 2000WO-US26524.  
 XX  
 PR 29-SEP-1999; 99US-0157137.  
 PR 03-NOV-1999; 99US-0163280.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
 XX  
 DR WPI; 2001-235357/24.  
 DR N-PSDB; AAH33488.  
 XX  
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 useful for preventing, diagnosing and/or treating colorectal cancers -  
 PS Claim 11; Page 6606-6607; 9803pp; English.  
 XX  
 AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patient's own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated P,  
 CC by inserting the nucleic acid into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAB77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX

SQ Sequence 178 AA;

Query Match 28.6%; Score 485; DB 22; Length 178;

Best Local Similarity 88.2%; Pred. No. 3.1e-42;

Matches 97; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

QY 36 FGMNPFVPLPPLAGLNOOIFFLCTTALAFLVMPVLRGNLLFRSLSSWFWL 95
DB 27 FCMGFSAYOTRFTCLGLVVOOILFFGLTALAPLVMPVLRGNLLFRSLSSWFWL 86
QY 96 TLAAVILONMAAHVWVLETHDGHQPLTRRVLVAATFLPLNVLVGAM 145
DB 87 TLAAVILONMAAHVWVLETHDGHQPLTRRVLVAATFLPLNVLVGAM 136

```

Search completed: March 13, 2003, 11:28:54  
 Job time : 42.9351 secs

